

OY	1029	catcgaggagaataatgctcaaaaagtgtaataattgcacaaatacgcagttctcttttgatg	1088
Db	1021	CATTGGAGGAAATGCTTAAACGGTTATTAATTGGCAAAATTACGGCATTTCTTTGTATG	1080
OY	1089	agactctaaagtacacttcagttctgcagatcagtcacaaacatgtagaataactcoccag	1148
Db	1081	AGACTCTAAGTACACTTCACTTTGGCCAGTACTGCCAAACATGTGAAATTACTCCCATG	1140
OY	1149	ctaatgsgtctcgtgagatgaaagcgtgcataaaaagtcacaaaggaataactbtygaat	1208
Db	1141	TTAATGAGGTCCTGGATGATGAAAGCCTTGCTTAAAAAGTACGAAAGAAATCTTGCAAT	1200
OY	1209	taagaagaacatttagagaattttagatcgtctgtaaacaaacctaagcaatgta	1268
Db	1201	TAAAGAAACAAATTAGAAATTAAGACTCATCGTCTTAACAAACCTCAAGCAATGGCTA	1260
OY	1269	aagaagaagcatalacagatttgcagtcgaaatcaaaacatcacaaagagaagaagta	1328
Db	1261	AAGAAGAGCATACACAGTTGCTGATCGAATCAACAACTACCAAGAGAGAGATA	1320
OY	1329	gaataatgacattgcacaataatgttgtgtctcatccagaatctcaacagagccaa	1388
Db	1321	GAATATGGCACTTGCAAAATATTGTTGCTTCAATCCCAAAATCTCAACAGGCCAA	1380
OY	1389	gggtcaagaagaaacggaaggttactgtggggccgaagaaataaccaaaatggttaacg	1448
Db	1381	GGGTCAAAACGAAACGAAAGTTACTGGGGCCGCAAGAAATTCGAAATTAATTACTATG	1440
OY	1449	ctctcgtgcttcctgbaacttgcataatgctaaccagatctactgcaattttagcaagaag	1508
Db	1441	CTTCTGGTGTTCCTGACTTTGATATGCTTATCCAGATTACTGGCAATTTTAGCAAGAAG	1500
OY	1509	caaaagttctctgacatgacctcaatttcgaagaattgtagacctgtttgtacagagttt	1568
Db	1501	CAAAAGTTCTCTGCACATCCCTTCATTTCCGAAATGTGATACCTCTTTGTCCAGAGTTT	1560
OY	1569	ctgaatttctgtagagccctccatcatgatatgacagcaatgaaataagatgcagaatgaaat	1628
Db	1561	CTGATTTTGAATGACGCCCTCTCCATGATGGACAGCAATGAATGATGACAGATGGAAAT	1620
OY	1629	tagccagataaagttaactcacagagaagaaagacatcacctcaatcaatcatgataagcttg	1688
Db	1621	TAGCCAGATAAAGTAACACACAGAGAAACATCACTTCATCAATCAATGATAGACTTGG	1680
OY	1689	gacagactctgcagagtggttcagttctcatgactctctaaagaaacacagctatacaatcc	1748
Db	1681	GACAGATTTTCTGACAGTGTTCACTTTCAATGATTTCTTAAGSAAACACAGCTACAAATAC	1740
OY	1749	tcaccaagaacctcgtgtagatagctgaaatgcagaagaagctctcttggaaaagagatca	1808
Db	1741	TCCCCAAGAAGCTCTGGTGATATGGCTGAAATGCAGAAACCTCTTTGMAAAAGAGATCA	1800
OY	1809	caaacctccagagcaaacactcagtcbaaagagagaaagaaagaaagaaacttacaagact	1868
Db	1801	CAAACTCCAGCAACACATCAAGCTCAAAAGAGAGAACAAAAAGAAAGAACTGTGCAAACT	1860
OY	1869	tcagagctcaagatagcagaacttggaaagacagcttagtgcataaagcttaaaaactctagaga	1928
Db	1861	TCGAGCTCAAGATTAAGCAGAACCTGGAAGACAGCTTAAAGTGTCAAACTTAAATGTAGGA	1920
OY	1929	tggtttacaaactcggagagagcatctcataatagctggaagtcacaaacagatgttgaagaag	1988
Db	1921	TGGTTAAACAACTCGAGAGACATTCATTAAATGCTGAAATCCAAACAATGTGTGAAAGG	1980
OY	1989	aagttctgcagaaagaagaatctcagtcctctggagagctcgtgttacaagtcatcaaacagtg	2048
Db	1981	AAGTTGTGAGAAAAAGAAATGTCATGCTCTTGAGAGACTCTGTGTACAAATGATCAAAACATG	2040
OY	2049	acctacagagatgctctgtttagtgaataacgtctaaagcagctcccaatgagtgatgataag	2108
Db	2041	ACCTACAGAGATAGTTGTTGTTAGTGAATGAACGCTTAAGCAGCTCCCATATATAGTGTATAG	2100

QY	2109	aaacacgaaaatctctgaaacaaagaatctgaatcttaaaagatcttatgaaaccta	2168
Db	2101	AACACAGAAAATATCTCGAACCAAAAGATCGTTATTTACAAAGTTTATTGAAAACCTTA	2160
QY	2169	acaagaanaagtgaaatcgataaaacaaaacctctctgacgaagatcttaatgaaagatctc	2238
Db	2161	ACAAGAAAAGCAGACATGATTAACCAAAAATCTTCTGACGAAGTTTTATGGAAGCTATTTC	2220
QY	2229	agctaattgaaagctataatgagcagaanaagcaaatgacatggaaggaaatctgacattatgc	2288
Db	2221	AGCTATGTGAAGCTATTAATGAGCAAAAAGCAAAATGCACGTGGAGGAATCGGCATTTATGC	2280
QY	2289	gagataatcttgaaacataatctctgagaatgaaaaccttaaaagggaatctgacatc	2348
Db	2281	GAGATAATTTTGACAAATATTATTCATGAGATTAACCTTAAAAAGGMAATTGCAGATC	2340
QY	2349	tggaacgtctcaatgaaggaataccaagaaccaaagagcttgaaattcctgaggaaggaat	2408
Db	2341	TGGAACGTTCTCTAAGAGATTCANMAACCATAGTTTGAAATTTCTGAGGAAGGAA	2400
QY	2409	ctcaaaaagaacacagagcacaactaaatccatbagaattgacagatctaaagaanaattagctg	2468
Db	2401	CTCAAAAAGAACAGAGAGCAACACTAATCATGATTTGGCAGTTTAAAGAAATTAAGTTG	2460
QY	2469	aaaatgcagagatgatacatcaaaatctctgaagaagactagaacataaacaacaaatctc	2528
Db	2461	AAAATGCGAGATGTATACAATCAAAAATCTTGAGAAATCTAAGAACTAAACAAACATTC	2520
QY	2529	tgaaaagacaggaataatccaactctgcagaaattaaagaaacgcgcagataactctgcagaaa	2588
Db	2521	TGAAAAGACACAGAAATTCATCTGCAGAAATTAAGAAACCCACATTAATCTTGCGAGAAA	2580
QY	2589	aagtaacgaatcttgatctctcgatcttccatggggaatagtgagaanaacctctgagaana	2648
Db	2581	AAGTACGAAATTTTGATCTCTCGGTTTCCATGGGTATAGTGAACACTCTGTGAAGAA	2640
QY	2649	tccttcaactgaagcaatctctctctctgagtctgaagcctgagacgcgcgcagatctcagaag	2708
Db	2641	TCTTTCAACTGAACCAATCTCTTTGTGATGCTGAACCTGTGACTGCGCATGCTCGAAGG	2700
QY	2709	aatgtctctctctcaagaagtgaaaatctctgagcttgaaaggaagaaatgtaggaacacatac	2768
Db	2701	AATGTCTTTCTCTCAGAAAGTAAATCTTAAGCTGAAGGAAATATGAGGACACATCTAA	2760
QY	2769	actgttaacaatcaaaagaagaagcctgcgcctctctgttgagaagcagctcgaagaatcgaa	2838
Db	2761	ACTGTACAAATCAAAAAGGAAAGGCTGCGCTTTGTTTGAAGACAGCTGGMAACTGAA	2820
QY	2839	aatcaaacataagaanaatggaagctgattctgcagaagaagatctgcaaaagtgccttaatg	2888
Db	2821	AATCAAACTACAAAGAAATGGAAGCTGATTTGCAGAAAGATTGCAAAAGTCTTTAATG	2880
QY	2889	agatctaactacttaaatgagcctctctgacaggaagatctccacagaaatcttgctctctctg	2948
Db	2881	AGATTAACCTACTTTAAATGGGCTTCTGGCAGGAAAGGCTCCCGACGAATTTGCTTCTCGT	2940
QY	2949	ctgaaatagaagaanaaggtctctgagatctctcaagaagcattgaagaacaaatctgagaagaa	3008
Db	2941	TTGAAATTAGAAGAAAAGGTTTCTGAGTTCTCAAGACGCTTGAGAAAGCATTTGAGAGAA	3000
QY	3009	aaaatgcctctgagaatcgaaagtgacttgcctcatcaagaatatacaaatctctgcacaaatcgaa	3068
Db	3001	AAAATGCTTGAGAAATGGAAGTGACTTGCTGCTATCAGAAATTAATTTTGGCAAAATGAG	3060
QY	3069	ctgaaatgcctgaaaaatcagaatcaagcaaggctctgaaagagataaatgcttaataagcaag	3128
Db	3061	TTTGATGCTTGAAAAAATCAGATCAGCAAGGCTCTGMAAGATTAATGTTATTAAAGCGAG	3120
QY	3129	aaggaagaacatctgcctcatatataagaacaaagaagatctcaatgcaaggagcaatctg	3188
Db	3121	AAGGAGAACATTTGCAATCTATTATTAAGCAAAACAAAGATTTATCTGCGAGGACAAATCG	3180
QY	3189	agcagatcttacaactctacgtgaagatgagcaacacaaagctcaaaagtgcagcagactg	3248

|||||
Db 3181 AGCACAATTTTAACTGACTGACGAGTACACACACACAGTCAAAATGCGACGACGACTG 3240
OY 3249 aagagcaactctggagatgaaagaaatgcatgcatcttttgaagaaatatacagaa 3308
Db 3241 AAGACCAATACTTGGAGATGAGAAATGCATGATGATCTTTTGAAGATATACAGAA 3300
OY 3309 acaaaagtgaagctgaagacctttaaagaaatggaagaaaccttaagcactagagat 3368
Db 3301 ACAAAAGTGAACCTGAGAGCTTTTAAAGGAATGAGAACCTTAAAGGCATATGAGT 3360
OY 3369 cgtggaatgaagattgtgcacacaaacatgaacttgaagaaacttaagcactaagga 3428
Db 3361 CTGTGGAATTAAGATTGCTGACACAAAACATGAACTTGAAACATNTAAAGGTAAGG 3420
OY 3429 agcaactgctcatgagaaaataacttttcaagcaatgcaagatlatatccgatla 3488
Db 3421 AGCAACTGCTCATGAGAAAATACTTTTCAAGCAATGCAGACTATATTTCCGATTA 3480
OY 3489 caactcttcagactcgctcctccctcaaaatagttgaaggaagcttcaagcccca 3548
Db 3481 CACCTCTTTCAGACTCGCTCTCCCTCAAAATTAAGTTGAAGGAACTCTCAAGACCCA 3540
OY 3549 tagaaatcaatgactacacaaatlaatatgacctgtctacagaaaggaacaaatlatg 3608
Db 3541 TGAATATCAATGACTACCAATTTAATAGCCCTTGCTACAGAAAGAACACATATGCG 3600
OY 3609 tgtgtctagagactgaagaaacagctctcaaggaagcaatlatatgattgaacactcaac 3668
Db 3601 TGTGTCTAGAGACTGAGAAACAGCTCTCAGAGAGCAAGTTATTTGAACACCTCAAC 3660
OY 3669 ttcaagttctcaagcaagaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaa 3728
Db 3661 TTCAAAAGTCTTCAAGACCAAAAGTATTTGAAAGTCTATCTCCAGAGCCAAAGCAAGT 3720
OY 3729 tggaaagaaagagagttaaactgtcttggagatggaaactgaaactgaaactgaaactg 3788
Db 3721 TGGAAAGAGAGAGGTTAAACCTTTTGAGATGGAACACTGAAAGGACACCTAACTG 3780
OY 3789 actcaagctgtctatagagaaatgcaagctggaagaaatcgtgaagcttcaagaaactcc 3848
Db 3781 ACTCAAGAGCTGTATAGAAATTCAGAGTGAAGTCTGGAAGTTACAGAAAACCTCC 3840
OY 3849 aaacacttcaagaaagatgaaatattactatagaaagaaatgaaatgaaatgaaatgaa 3908
Db 3841 AAACACTTCAAGAGAGATTAATTTACTATAGAAAGAAATGAGCTTCAACCAACT 3900
OY 3909 tcaagaaactgaagcagagcatgtagccttaaaacaaagccttaagtgaaacattgagc 3968
Db 3901 TCGAAGAGCTGAAAGGAGCATGATGACTTAAACAAAGACCTTAAAGAAACATTGAGC 3960
OY 3969 agtcaatgaaacaaagatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaa 4028
Db 3961 AGTCAATTTGAAGAACAGATGAATTAAGGGCTCCCAAGAGCTTAAGAGAACAGAGC 4020
OY 4029 aactggttgaacttagacaacagcctttagatgtctgttagaagaaatcacaacaa 4088
Db 4021 AACTGTTGATGACTTAAACACAGCTTTAGATTTGTTCTGTAAGAAATTTCAACCAA 4080
OY 4089 accatgtagcagttgcttaacaaagaaagggtgcaatgggtggaagaaatgcaatgcaatgca 4148
Db 4081 ACCATGATGCAAGTGTAAACAGAGAAAGGTGTCATGGGTGAAGTAAATTCCTTCAAA 4140
OY 4149 gttgaatgctgctgtgtggaagaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaa 4208
Db 4141 GTGAAGTGTGCTGCTGGAAGAGATGAGCTTCAACATCTTTGTAAGCAATTTAGTTTCAAG 4200
OY 4209 aactggaactactcgtgctcatgtaaaatctgtggaaggaagaaacacttgaatacaca 4268
Db 4201 AACTGGAGACTACTGCTGCTCATGTAAATCTGTGAGGGGAAACCTTGAATACAA 4260
OY 4269 aaaaactcaatgacctgaaagaaatattgggcaaatctgaaagaaaggaaggtgttga 4328
|||||

Db 4261 AAAAATCAATGSCCTTGAAAGAGAGATATTGGCAAAATCTGAAGAAAGCGAAGTTCGA 4320
OY 4329 agtccatggttggagaaatctaaagaaagacaacaataagcctcaagaaagcaagcaggaat 4388
Db 4321 AGTCAATGTTGGAATCTTAAGAGACACACATTAAGCTCAAAAGAAACAGAGAGAAAT 4380
OY 4389 attcagtaaaagaaatcaatcagtttagaagaggtgttcaagttgttcaagaaactgtg 4448
Db 4381 ATTCTAGTAAGAAATCAATTCAGTTTAGAGAGGTGTTCAAGGTTCACAGAACTCG 4440
OY 4449 tagatgagatagagctctgaaagcacaagcctaaaggaagcaagaaagcctgaaataa 4508
Db 4441 TAGATGAGATGAGCTCCTGAAGCACAAGCTTAAAGGCGAGAGAAAGCGCTGCAAAATTA 4500
OY 4509 aggaatagatatttgaacttgaacttgaacttgaacttgaacttgaacttgaacttgaacttga 4568
Db 4501 AGGATAGAGATTTATTTGACTTGTACAAATGCAAAACCAATTTAGTTGAGGAAAT 4560
OY 4569 tggaaactcacttgaagctgacaaatgaaagaaagaaagaaagaaagaaagaaagaaagaa 4628
Db 4561 TGGAAACTCCATTCAGAGCTGACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4620
OY 4629 tggagataaagcttctggagaaagaaacttggagaaagaaatcaglatcttgaagaaatctg 4688
Db 4621 TGGAGATTAAGGTTCTTGAGAAATCTTGAGCGAATCAATATTACTGCAAAAGATTGC 4680
OY 4689 aagaagaaagctggaactgtctaaacaaacttgaatccttcagaaagaaatgagaagct 4748
Db 4681 AAGAAGAAAGCTGGAACCTGTAAACAACTTGAATCTCTCAAGAAAGAGATGAGAGCT 4740
OY 4749 ccgttctatlaaagaatgaactgtcaacaaagcctagaagaaagcttgaagaaagaaatla 4808
Db 4741 CCGTCTATTTAAAGATACCTGCAACAGAAAGCTGAAGAGCTGCTGCTGTAAGAAACATTA 4800
OY 4809 ttctaaagagaatatgtacaacaaacctaaagcattcagataactgaactgaactgaactgac 4868
Db 4801 TTCTTAAGAGATTTTGAACACACCTTAAGCAATCAATCAATCAATCAATCAATCAATCAAT 4860
OY 4869 agaaacacagcaagagctacagttgtgctaaagaaatcgtcaatcgtctctcaacaaat 4928
Db 4861 AGAAACACAGCAAGAGCTACAGTTGGCTTAAGAACTTGAACATGCTCTCTCAACAAAT 4920
OY 4929 gtccaataactaaagaaagaaactcttgaagattgtgtgtcatctcttgaagaaagaa 4988
Db 4921 GTCCATATACCTCAAGAAAGAAAGCTCTGCAAGATTGTGTATCTCTGCAAGAAAGAA 4980
OY 4989 tatgttataactgaagaaatgcaatcaaaactaaatgaacaaagaaatattactacatg 5048
Db 4981 TATTGTTATTAATCTGAAGATTTGCATCAAAACCTTAATGAACAGAAATTTACTACATG 5040
OY 5049 aaaaagaaatgaaactgaaacagctcaagttgagcctaaagtgagagtggaacatctgata 5108
Db 5041 AAAAGATGAACCTGAACAGCTCAAGTTGAGCTTAAAGCTCAAGCTGGAACATCTGATGA 5100
OY 5109 aagatgattgaaatcgaagttcctcaacttgaagttccttgaagaaatgaaagaaatgaaatgaa 5168
Db 5101 AAGATGATGCAAGTGAAGAGCTCTCACTTGAAGTCTCTTGCAGATGAGAAACATGATGACTG 5160
OY 5169 aacaaacttctgtctttaaagcagatgcaagtagttagttagttagttagttagttagttagttag 5228
Db 5161 AACCAACATTTCTGTGCTTAACAGCAGATGCAAGTACTCTCAAGAAAGAAAGAGAGC 5220
OY 5229 tgaacaaacccaatgaacacttaacagctgaggtgagcacttaaaagaaatgaaatgaaatgaa 5288
Db 5221 TGCACAAACCCATGAACTTAACAGCTGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 5280
OY 5289 tgggttgaattttaaaatgaaagcgaagaaagaaagaaagaaagaaagaaagaaagaaagaaag 5348
Db 5281 TGGGTTTGAATTTTAAATTTGAAGCCGACCAAAAGCAACCTAAAGAGCAATGTCTGTAA 5340
OY 5349 atgagataaagaaacttgaagcaagcagacagacttaaatgtgagatagaagaaactta 5408
Db 5341 ATGAGATTAAGAACTTTGAGCAAGCCACAGACTTCAATGTGATGATGATGATGATGATGATG 5400
|||||


```
|||||
Db 7561 TCTTAACAAAAAGCTAAAGGAGTGTGTACGTGCAAAATTAACGTAGATTAACCCGTAT 7620
OY 7629 atctgaaccaggttcgaagcgaatctgcaagaagaagaacaaacaaagaacttatgc 7688
Db 7621 ATCTGAACCAAGTGTGAAGCAATATGCAAGAGAAAGAAACAAACAAAGACTTATGC 7680
OY 7689 gcaagaatgagcaacacagccctagtgctagtgtaatgaggaagaagaatgctagactt 7748
Db 7681 GCGAATGAGACACACGCGCTAGTGTATGATGAGAGAAAGAAATGCTAGACTTT 7740
OY 7749 taggcactgtaaaacttcaagatgaatccaaagaactccaaatcaagatcaaatgc 7808
Db 7741 TAGGCATACTGAAACTGTTCAAGATGATATCAAGAAACTCAAGATCAAAATGC 7800
OY 7809 tagaataatgaactgaacttgg tcaagaatgcatgacacaaagtgaagaatgtgcga 7868
Db 7801 TAGAAATGAACGTGAATGTTGCTCAAGATGATGCCATGCAAAAGGTAAGAAAGTTGCAA 7860
OY 7869 ttctgaagacaaactactaagcagaacgcggaagctgaagcctaaacgcaatgcaqtga 7928
Db 7861 TTTTGCAAGACAACTACTAAGCAAGAAACCGCAAGCTGAGCTAAACGCAATGCAAGTGA 7920
OY 7929 aactaactaaagaacgaataatctcaagctgcaatgcaatgaagaataagaaactacaga 7988
Db 7921 AACTAACTAAAGCAAGATTAATCTTCAAGCTGCATGAAAGAAATAGAAACCTACAGAGA 7980
OY 7989 aaatggtgtccaaagtgtgacgtaccataaagaagaatltgacaacttaaaactaagg 8048
Db 7981 AAATGTTGTCGAAGGTCAGTACATATTAAGAAAGAAATTTGACAACTTAAACTTAAGG 8040
OY 8049 tggtaagaattgaatgaaaaataaagtaactcaaaagacacagcaagaagatgcct 8108
Db 8041 TGTATAAGATTGAATGAAAAATTAAGTACTCAAAAGCAACGACCAAGATTTCCCT 8100
OY 8109 acttaagctctgttttggaagaagaagaagcttgcgtataggttlaaagaagaactta 8168
Db 8101 ACTTAAGTCTGTTTGGAGATTAAGAAAGGCTTCCGTAAAGAGACTTA 8160
OY 8169 ggcgagcacagcagacacagatacaacagtttgttccaaagatltacagaagactt 8228
Db 8161 GCGGAGCACAGGCAACAGATACACAGTGTGTTCCAAAGATTAACAAAGACTT 8220
OY 8229 caacttccctgtactgtgtgtgtgtgaagcggatagtgacagacacagcaatcttg 8288
Db 8221 CAACTTTCCTGTGACTTGTGTGTGTGAAGCGTATAGTGCAGACACAGCAATCTTG 8280
OY 8289 tgcgtcagtcggaacaaagccgcttgaagaaggagctgttcaacttaagaagaataltc 8348
Db 8281 TGTGTGAGTGGAAAAAGCGCGCTTGGAAAGGAGCGCTGTACATTTAAAGAAATATATC 8340
OY 8349 atcatattacaagactatgtaagttcgaagatcgaagaagaacaaagcaaatctg 8408
Db 8341 ATCATTTATACGAGACATATGTCAAGTTCTGAAGATCGAAAGAAACAAAGGCAAAATCTG 8400
OY 8409 atctcaattcttcaactgatactcacacagagcctcgctccaaactgaactt 8468
Db 8401 ATCTCATTTCTTCTCACTGTGATCATCACAGAGGCTCCGCTTACAAACTGAACCTT 8460
OY 8469 acagacatgcccctgttacttccagaagaagttcgaatgtccaaagcctcaacttagatctc 8528
Db 8461 ACAGACATGCGCCTGTTACTCCAGAAAGGTGTGAATATCCCAAGCTACACTAGATCTC 8520
OY 8529 caaagaagtcggaagtcagacactaaagttgtgtgtccaaagaagaagttcgaagaatctaca 8588
Db 8521 CAAGAAGTCGGAGTCAGCACTAAACGTGTGTGTCAACCAACAGAGTCCGAATATGTACA 8580
OY 8589 gccaatagtaatgtctcagaagccggaatgataacataacttctccaaagca 8648
Db 8581 GCCAATTAGTATGTCTTCAGAGCAAGACCGGATGCACTAAACATATATCTTCCAGACA 8640
OY 8649 aggttgagctgcacaaaagctgctgtctccaaacataatgagatgcccacccagc 8708
|||||
```

```
Db 8641 AGTTGAGCTGCACAAAAAGCGTGCCCTGTCTCCAAACAGATGAGATGCCACCAC 8700
OY 8709 atgtcatacccttgcgaagccgactgcatataaaatcctaactgaagcagttatcg 8768
Db 8701 ATGTCAATTCCTGTGGCAAGCCGAGCTGATTAATAATCTAACTGAAGACAGTTATTCG 8760
OY 8769 acaattgtcttctccatgcaaacagcaaaaagtaacaggaatcctaattccctaaag 8828
Db 8761 ACAATTGTCTTCTCCATGCAACAGCAAAAAAGTACAGGAATATCAATTTCCCTTAAG 8820
OY 8829 gcaaatatgtatgtaagttcaaaatcgatgccttaactgcctacatcagtttltgata 8888
Db 8821 GCAATATTATTGATGATGAATCAAAATCAATGCTTACATGCTCATTCAGTTTGTGATA 8880
OY 8889 atttaagcttggatgtatttccagagctcaacacagcagaagagcaatgaagaatgcagg 8948
Db 8881 ATTCTAAGCTTGTGTGATTTTTCAGAGCTCACACAGCAGAGCAATGCAAAAGTACAG 8940
OY 8949 ctgagaactgtgtgtatgtagcaaaaaaagaagcgcaactgaaatgtataaacatcctaga 9008
Db 8941 CTGAGAACTGTGTGTATGAAAGCAAAAAAGAAACGCACTGAATGTAAACATCTAGA 9000
OY 9009 tccctgtacatctgactctcctgtctcgaagaagactgtgactctgtcttcttagg 9068
Db 9001 TCCCTGTACATCTGACTCTCCTGCTCCCAAGAGACTGTACTCTGCTTCTGTAGG 9060
OY 9069 aagaacacatgaacactgcactgtctcgaataaagaagtcactcagtgagaacaaagtgtt 9128
Db 9061 AAGAAACACTGAACACTGCCATGTCTGTGATTAAGAAAGACTCTCAGTGAAAGAAAGTTGTT 9120
OY 9129 ctcttgatgtaactcagtggttggcgagtggttaagctcttaaaatgaagtcaatagt 9188
Db 9121 CTTAAGTATATACACGTGTGGCGAGTGTTACGCTTTAAATAAAGTCAATATAGT 9180
OY 9189 ctcaagctatattatagtattagtcgtgtattgttcaacttlaagtccttgact 9248
Db 9181 CTCAGTCTATTTATATGTTATGTATGTGTATTTGTACCTTTAAGTCCCTGACTT 9240
OY 9249 catatttggctcactcgttagttcttctgtgtgtgcacacacacacggtggaataatga 9308
Db 9241 CATATTTGGCTCATCTGTATGTTCTTGTGTGTGGCACACACACAGTGTGAATAATGA 9300
OY 9309 catltagcgtatgttatttactgactgtctcctcgtgggcacatccatgacaaacacat 9368
Db 9301 CATTTGAGAGTATTTTATTACTGACTGCTCTCTGTGGGCCATTCATGCAAAGCACCAT 9360
OY 9369 taagtgtccaaatgttlttaactactatattatgtctgacttgtgaataagaataatc 9428
Db 9361 TAGTGTCCAAATGTTTTCACACTTATTTATATATGTGTGACTTTGTGAATAATGAATATC 9420
OY 9429 tacaagaatagggcaaaagtgttcttccactaagttaaacagatagggttlaactaat 9488
Db 9421 TACMAAAGATAGGCAAAAGTTGTTTCTCACTAGGTAACGATATGGGTTTAACCTTAAT 9480
OY 9489 tatttctcaataaaatattctataataatgagcaltcaattagggtgacctaagttatcc 9548
Db 9481 TATTTCTTAATAAATATCTATATATGATGCAATCATTAGGTGACCTAAATGTTATTC 9540
OY 9549 tgcactctgttaaaccttlttaattcaactcgttaataaagaagccctgatttlaaatc 9608
Db 9541 TGCACTCTTGAACCTTTTAAATTTCACTGTATATTAAGACGCCCTGATTTAAAT 9600
OY 9609 aaaaaaaa 9618
Db 9601 AAAAAAAA 9610
```

```
RESULT 2
HSCENPE
LOCUS HSCENPE 8257 bp mRNA linear PRI 10-JAN-1993
DEFINITION H.sapiens CENP-E mRNA.
ACCESSION Z15005.1
VERSION Z15005.1 GI:29864
```


QY 1023 aaattcattggaagaaatctaaacggtataatttcgaacttaagcttctt 1082
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 968 AGAATTCCTTGGGAGGAATCAAGACGATATATCTGCACATTTACTCCAGATCTT 1027
QY 1083 ttgatgagacttaagtaacttaagcttgcgaactgccaacatgltgagaatctc 1142
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1028 TTGATGAACACTTCTACTGCTCTCGATTGGCAGTACTGTAATATATGAAAGTACTC 1087
QY 1143 cccatgtaagtaggctctgtagatgtagaagctgtgtcctaaaggtacagaagaaatct 1202
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1088 CTTATGTAATGAGGTATCAACTGATGAGGCTCTCTGTAAGAGTATGAAAGAAATTA 1147
QY 1203 ttgatgtaagaaacaaatagaaatcttagatgacatgctcgtgaacaaagctcaagcaa 1262
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1148 TGGATCTTAAATAACATTAAGG-----GAGCTTCTTTAGAGCGCGGCTCAGGCA 1201
QY 1263 ttgctaaagaagagcatcacagttgttagctgaaatacaacaactcacacaagaagag 1322
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1202 TGGAAAAAGACCAATTGGCCCACTTTGGAAGAAAAAGATTGGCTCAGAAAGTACAG 1261
QY 1323 aagatagatgtagcactgtagcaaatatgtgtgtg---cttaccacaaatctcaac 1379
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1262 ATGAGAAATATGAAATTAACACGATGCTGTGACCTCTTCTCCTCAAGTTGCAAC 1321
QY 1380 aggcacaaagggtcaaacgaacgaagatgacgtggcgccgaagaaatccaaata 1439
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1322 AGGATTTAAAGGCTAAAGAAAGCAAGATGTTACTGTGCTTGCGCAAAATTAACAAA 1381
QY 1440 gttacatgctctgtgtgtcttctgacttgatgatactcaagattaccgtgaattta 1499
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1382 TGAAGA-----CTCAACATGATGCAATCAATTAATATACCAACAATATTA 1429
QY 1500 gcaagaagcaaggtctctgacatgaccttccagaaatgatactgtgtta 1559
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1430 CACAAAACACATPACGCTTCTATTAATTAATTAACGAAATGATGATCTGTGT 1489
QY 1560 cagaattctgatttgatgacgccct 1587
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1490 CAGAGTCTGATGTTTTCAGTAACACTCT 1517

RESULT 3
AB028468 2576 bp mRNA linear PLN 25-DEC-1999
LOCUS Arabidopsis thaliana ZCF125 mRNA for kinesin-like protein, complete cds.
DEFINITION
ACCESSION AB028468
VERSION AB028468.1 GI:6526974
KEYWORDS kinesin-like protein.
SOURCE Arabidopsis thaliana (strain:Columbia) cDNA to mRNA.
ORGANISM Arabidopsis thaliana
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
AUTHORS Kato, A., Suzuki, M., Kuwahara, A., Ooe, H., Higano-Inaba, K. and Komeda, Y.
TITLE Isolation and analysis of cDNA within a 300 kb arabidopsis thaliana genomic region located around the 100 map unit of chromosome 1
JOURNAL Gene 239 (2), 309-316 (1999)
MEDLINE 20018182
REFERENCE 2 (bases 1 to 2576)
AUTHORS Ooe, H., Kato, A. and Komeda, Y.
TITLE Direct Submission
JOURNAL Submitted (05-JUN-1999) Hidekazu Ooe, Graduate School of Science, Hokkaido University, Division of Biological Sciences; Kitaku Kita10 Nishi, Sapporo 060-0810, Japan
(E-mail: ooe@sci.hokudai.ac.jp, Tel:81-11-706-2743, Fax:81-11-706-4851)

FEATURES
source Location/Qualifiers
1..2576 /organism="Arabidopsis thaliana"

/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="chromosome 1"
/map="between m1303 and m1259"
8..2479
/gene="ZCF125"
8..2479
/gene="ZCF125"
/codon_start=1
/product="kinesin-like protein"
/protein_id="BA08112.1"
/db_xref="GI:6526975"
/translation="MEKICVAVRVRPAPENGASLWKVEDNRISLHNSLDTPPTTASH
AFPHVEDSTNSVYELIKDIIHAVEFNGEAFYAGTSSGKFTFMGSETPDI
IRSRVDFEIRIHSREFLIRVSYEIEINENFVNDSEFRNOSDEPIISRTPDGICVNDSDVPG
GKEEIVSDAEOLIKLIDSGEVRNHRFETNMVNSHSHIIPFVIESRDKSSDA
IRSVLWVLDLGSERLAKTGAGGVRLOBEKYLKSLMITGIVINKLSDBTKLRHP
YRSKLRIILQALGNAKTCIICTIAPBEHHIEESGTLQFASRAKRLITNCAOVNI
LTDAAILKRLLEIEELRMKLGSHAEVLEILNLSNOMLYELBEERLTKOLEBK
RKOEONCIKEOQMIENLNNEVNDSEFRNOSDEPIISRTPDGICVNDSDVPG
TPCKRSASRSPVARSNNYSGLSDSPSVSLGVDVEDIKRSLITLITNENSLQVKEOYLISNNIMO
PAYCOPTPISTATTECPRENHSPVEDIKRSLITLITNENSLQVKEOYLISNNIMO
EMSEKQETITVKEIPKRLSESVANCKDYIKDYIVTKSLITTKESPFLMLITTEI
TTSILATLETFQSMIDQKTTGSSIDHPLSDHETLKVNLKNTTLLSDAQKDEFL
NSHNKQETALEBKLLSELIIIEKRYNLEKELCDKQLASRSHKLIKEOVF
LKEBRSDLRKISQSTORLIVIASDKENALDLNVEYKRRKDEEETKHISIAFATRH
KSFVSPFSEIKSMOKLTTONSKAP"

BASE COUNT 856 a 461 c 589 g 670 t
ORIGIN

Query Match 2.5%; Score 243.2; DB 8; Length 2576;
Best Local Similarity 57.0%; Pred. No. 7.2e-37;
Matches 540; Conservative 0; Mismatches 378; Indels 30; Gaps 4;

QY 272 tcttcaattcgatcgigtatctaattcgaatcaacaaatcaattaccaga 331
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 134 TCTACGCTTTTGATCATGATGTTGACGAAAGTTCTACGAATGCTAGTCTACGAGCTT 193
QY 332 atagacagactcatcagatcagcttgagagataatgacaaatctgacatc 391
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 194 CTTACTAAGATATATACATGACGCGCTTTGAGGCTTTTAATGATCGCATTTGCTTAT 253
QY 392 gacagacatcttcaagcaagcgtacacaaatgatlgygaacacaaatcttgagcata 451
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 254 GGGCAACTACTAGGGAAGACGCTTACGATGACGGGCTCTGAACATGATCTGGAAT 313
QY 452 ataccacagccatacagaagttttaaatcttaagagatcacgaacagaagatt 511
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 314 ATTGGAACATCTGTGCGATGCTTTGAAAGAAATACATATGATATCGATCGAGCTT 373
QY 512 ctctaagatcttctaataatgagatttaacaatgaaactggaagactactgtagat 571
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 374 CTCATCCGCGTTTGTCATGCAATTTACACAGCAAGAAATTAATGATCTTTAGCTGT 433
QY 572 gacagaagaagacccttgaatctcgagagatttaataagaaacgtgtatgtgct 631
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 434 GA-----GACCAAGAGACTCAATTCATGAGCATTTGAGCGCTGGAGTGTGTTGCT 487
QY 632 gacttgactgaagaactgtatagttccctgaacaatgataacagtgatgataaagggt 691
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 488 GGTCTTAAGAAAGAAATTTAGATGCTGAACAGATTCATTAAGCTTAATGATTCGA 547
QY 692 gaaataaagacacattagagagacaaatgataatcatagtagtctgcatcacata 751
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 548 GAAGTTAATAGCACTTTGGCGAGACAAACATGATGCTCTATAGCAGGCTCTCACACC 607
QY 752 atattagaatgattgttgaagcagagacagaatgatacccaacaaatcagagaactgt 811
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 608 ATATTCAGAATGATGATGAGAGCGAGGAAACATAT-----AGTTCT 652
QY 812 gatgagctgcatgatactcaactgaattgtagatcttctgtagcagtgaaagca 871
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 653 TCAGATGCTATTCGTCTGCTAGCTTGAATTTGTCGATTTAGTCGATGAGCGAATT 712
OY 872 agccaaactgagctgaaggtgtgtgacttaaggaagctgacatcaaccgcagcttg 931
Db 713 GCTAAACTGGGCTGGTGTGTACGCTTGGCAAGAGGAACTGACATTTAAAGAGCTTA 772
OY 932 ttatcccttgacagagcttaataagacttagcgacg--ccagcggtgtgtatata 988
Db 773 ATGATCTTGGTAATGTCATCAATAATTAACTGACAGTACAAAGCTAAGGACATATT 832
OY 989 aactacagagacagaaactaccagaattctccaaattcaatgtggaggaatgtctaa 1048
Db 833 CCGTACCGAGTAGTAACCTGACTCGAATCTTTCAGCTGCGTGGTGGTAAATGCAAA 892
OY 1049 acggttaaatgtgacaaataacgcaag-----ttcttttgatgagacttaagtaca 1102
Db 893 ACTGCAATTAAATGACATATCGCACCAAGAGCATATTTAGAGAAATCAAAAGAACT 952
OY 1103 ctccagttgcagctacgcacaaacatgtgagaatactccccatgttaatgagctctg 1162
Db 953 CTCGAATTTGCAAGACAGCAAGCAAGCGCATCACCAACTGTGCTCAAGTGAATGATCTTG 1012
OY 1163 gatgataagcgtgtgtctaaagaagtacagaagaagaatcttggattta 1210
Db 1013 ACTGATGCCCTTATTTGAGGCGCAAAAGTTGAGATAGAAGAGCTA 1060

RESULT 4
AB001426 480 bp mRNA linear ROD 19-SEP-1997
DEFINITION Mus musculus mRNA for motor domain of KIF10, partial cds.
ACCESSION AB001426
VERSION AB001426.1 GI:2443267
KEYWORDS motor domain of KIF10.
SOURCE Mus musculus (strain:ICR) 4 week cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 480)
AUTHORS Nakagawa, T.
TITLE Direct Submission
JOURNAL Submitted (25-FEB-1997) Terunaga Nakagawa, University of Tokyo
Medical School, Department of Cell Biology and Anatomy, 7-3-1
Hongo, Bunkyo-ku, Tokyo 113, Japan
(E-mail:nakagawa@kinesin.kaiyodai.m.u-tokyo.ac.jp,
Tel:81-3-3812-2111, Fax:81-3-5689-4656)
2 (sites)
REFERENCE Nakagawa, T., Tanaka, Y., Matsuo, E., Kondo, S., Okada, Y., Noda, Y.,
AUTHORS Kanai, Y. and Hirokawa, N.
TITLE Identification and classification of 16 new kinesin superfamily
JOURNAL (KIF) proteins in mouse genome
MEDLINE Proc. Natl. Acad. Sci. U.S.A. 94 (18), 9654-9659 (1997)
FEATURES 97420736
source location/Qualifiers
1..480
/organism="Mus musculus"
/strain="ICR"
/db_xref="taxon:10090"
/dev_stage="4 week"
<1..>480
/codon_start=1
/product="motor domain of KIF10"
/protein_id="BAA22386.1"
/db_xref="GI:2443268"
/translation="IFAAGTASGKTHPTMGSEDLGVIPRAIHIDFQRIKPFPEREF
LRLVSEIYENETITDLCAKQKKPLIEDNTRVYVSDLFEVYVTAEMAKMLA
TGEKNRHYGTIKNNQSSRSHTIFRMILBSREKAPSPSCGSAVSVSHIKLIDYDLEGS

BASE COUNT 162 a 87 c 106 g 125 t
ORIGIN
Query Match 2.2% Score 209.2; DB 10; Length 480;

Best Local Similarity 67.9%; Pred. No. 3.4e-30;
Matches 292; Conservative 0; Mismatches 138; Indels 0; Gaps 0;
OY 380 aatttgcatacagacagacatctcaggaagacagctacacaaatgttggaaacccaat 439
Db 1 ATCTTTGGCTACGGCAAACTGCAATCAGCAAGACACACATGATGATGCTTCTGAAAT 60
OY 440 tcatgggacataatcccccaagccatcacagagcttcttaaatatcatcagagataccg 499
Db 61 TGTTTGGAGTTATACCCAGGGCAATTCATGACATTTTCCAAAGATAAAGATTTCCT 120
OY 500 aacagagatctctcctaagagttctctatataatgagatctacaaatgtgaaacgtcgaagac 559
Db 121 GAGAGAGATTCTCTCTAGAGTTCTTACATGAGAGATCTAATAAGAACCTTACAGAT 180
OY 560 ctactgtgtatgacagagaagaagacccttggaattcgcgagatcttaataagaac 619
Db 181 TTACTCTGTAAATGCACAAAATGAACCTTTGATTAATTCGGGAAGACACCAATAGGACT 240
OY 620 gtgtatgtgtcgtgacgtgacgtgaagaactgttaatgtgttcctgaacatgtatcagtg 679
Db 241 GTATATGTTTCTGATCTCACAGAGAGAGTGTATACAGCAGAAATGGCTTTGAATGG 300
OY 680 atcaaaagggtgaaaaaaacagacatataatgagagactaaatgaaatcatatagtagt 739
Db 301 CTCGCAACAGGAGAAAAAAGACACATTATGSAATACCAAAATGATCAAAAGAGCAGT 360
OY 740 cgttcacatacaatattagaatgattgttgaagcagagacagaagaatgataccacaat 799
Db 361 CATTCTCATACATTTTATGATGATCTGTAAGTGTGAAAGAAAGCGTACGCTTCTAAC 420
OY 800 tcaaggaact 809
Db 421 TCGATGGCT 430

RESULT 5
AX321334 581 bp DNA linear PAT 15-DEC-2001
LOCUS AX321334
DEFINITION Sequence 351 from Patent WO0177168.
ACCESSION AX321334
VERSION AX321334.1 GI:17905139
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Lodes, M.J., Wang, T., Mohamath, R. and Indrias, C.Y.
TITLE Compositions and methods for the therapy and diagnosis of lung
JOURNAL cancer
MEDLINE Patent: WO 0177168-A 351 18-OCT-2001;
FEATURES CORIXA CORPORATION (US)
source location/Qualifiers
1..581
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 205 a 101 c 130 g 145 t
ORIGIN
Query Match 2.1% Score 206.6; DB 6; Length 581;
Best Local Similarity 68.4%; Pred. No. 1.1e-29;
Matches 333; Conservative 0; Mismatches 144; Indels 10; Gaps 3;

OY 955 gaagcttagcagcagcagcgtgtgtgataataactacagagacagcaaacacacag 1014
Db 72 GAACTTAGTGATGACAAAGTTGGTGTTCATTAATTAATGAGATACCAAGTTAACAG 131
OY 1015 aattctccaaatcatctggagagaaatgctaaacggttaataatctgcacaaatcagcc 1074
Db 132 AATTCACGAATTCCTTGGGAGGAAATGCAAAAGACAGCTATTATCTCACAATATCTCC 191

Query Match	Best local similarity	Match 464	Conservative	Mismatches	Indels	Gaps
360	1.8%; Score 177.8; DB 3; Length 7524;					
548	53.0%; Pred. No. 2.8e-24;					
420	0;					
608	387;					
480	24;					
668	3;					
540						
728						
600						
782						
660						
839						
720						

Db 899 ACATGAACGACGCTCCAGCCGTCGATGCATATTTAAGATATCATGAGCTCCGAA 958
 QY 780 acagaataatccccacaattcagaactgtgagtcgtcatgtatctcacttga 839
 Db 959 AGTCGGA-----CCACAGCGATGACGACGACGATGATTCAGAGCGTTTGA 1003
 QY 840 atttgtagatctctgtcgtcagtgaaagagaagcccaaatctgtagagtgtagagc 899
 Db 1004 ACCTGGTGAATTTGGCTGGCTCAGACCGGGCGGACCAAAAGGGTCTCGAGGAGCGGCC 1063
 QY 900 ttaagaagagctgcacaacatccagctgttatacctctgacaggtattataagaagc 959
 Db 1064 TCAAGAGGGCGGTCTATTAACAAGAGTTACTTTTCTCATACGTTATCAAAAGTC 1123
 QY 960 ttagcagcgagcagctgtgtgattataaactacagagaagcaaacctcacaagaattc 1019
 Db 1124 TCTCGGAAAAACCGGACATAGATTATACCAACACCGCATTCACAGCTGACCGCATTC 1183
 QY 1020 tcaaaatcatcttgaggagaatgtcctaaaggttataattgacaaattagccagttc 1079
 Db 1184 TCGAGGCTTCTTTGGCGGCAATGATTCATCATTCATTCATCAATCAATCAATCAATCA 1243
 QY 1080 ctcttgatgagactcgaactcagctcagcttgacagctcagcaacatgtgagaata 1139
 Db 1244 TCATGGAAGAGTGCAGCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCT 1303
 QY 1140 ctcaccatgttaatgagtcgtcgtgagtgagcgttgcctaaagagtcagaagaagaa 1199
 Db 1304 ACCCCCAATCAACGAGATGATGATTCGACGCCACATGATGATGATGATGATGATGATG 1363
 QY 1200 tcttgatttaagaagaactagagaattagag 1234
 Db 1364 TAAAGGCTTAAGATAGCTGCGCGAAGAAGAG 1398

RESULT 7

AB062739

LOCUS AB062739 3692 bp mRNA linear PLN 23-NOV-2001
 DEFINITION Arabidopsis thaliana MKRP2 mRNA for kinesin-related protein, complete cds.

AB062739

ACCESSION AB062739.1 GI:16902293
 KEYWORDS Arabidopsis thaliana cDNA to mRNA.

SOURCE

ORGANISM

Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

Itoh, R., Fujikawa, M. and Yoshida, S.
 Kinesin-related proteins with a mitochondrial targeting signal
 Plant Physiol. 127 (3), 724-726 (2001)

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..3692

/organism="Arabidopsis thaliana"

/db_xref="taxon:3702"

301..3468

/gene="MKRP2"

301..3468

/gene="MKRP2"

/codon_start=1

/product="Kinesin-related protein"

/protein_id="BAB71852.1"

/db_xref="GI:16902294"

/translation="MASSSRTSSRPPSPASSSTSSSHLSNLNIPRNSSTSSSLITS"

AAGIASRMSPTSFSDSGLIGSGFSGISPPVPSSELLDPMDDTISSESDISIST
 VRRPRLSDREYORGDENVAMYPDGDITLHREYVPLTAANDKXVGEQAATTIDYDAAR
 PIVKAAEMEVNTVAIVGYTSSGKTLTHMGDDESGITPLAKDFPSITQDPGPEPL
 LRVSTLEIYNEVINDLDPGONLRREDISGTYVEGKEEVVLSGHALSTIAGEE
 HRHVGSNNFNLLSSRSHITFTLGMSEKATHDYDVITFSQLMLDLAGSESSKFTTG
 LRKKGSGYINKSLITLGTITLGMSEKATHDYDVITFSQLMLDLAGSESSKFTTG
 TPASSSEETNHTLKFAPRAKSIETIYASNOIIDEKSLKTKYRETSITKLELDLR
 GMLVGSHEELMSTLKQOLEGOVKMSRLFEEREAALMSRIKTLKTLIVSTKNSI
 PGYSGDIPHTQSLNSGKDKDFDSLSESDNGSPSTLALSESLGPNHRSSSKL
 NDENSPGAETGTGVMPDEIDLLVEQVKMLAGEIAPSTSLRLVDQSVNDENSO
 IONLEREIHEKOROMGLEOLIIESSGASADYDVITFSQLMLDLAGSESSKFTTG
 KSNADNIOLOEOKTEENKLEHRYNLLLEQRLNVSSEKSPSCSNKASVGEVDEL
 KKIOSOIEENELKLEHRYNLLLEQRLNVSSEKSPSCSNKASVGEVDEL
 EYTKLSIOWTKLEKELAAARDLAOTRNPNVNGYRRKRYNCAKBSRGRTSSSSGDER
 DAMNIDPEDLKLMEVLRKOREVALSALAEKCTFIDEYTRKKAERKREALENDLAN
 MNVLAUKLKKDQALPEPNTDGRLEKESQHAVALKEQVSSAPRPPVVAATTEE
 TPKEPELVARLAKRMQEMKEKEMKQANDANSIHKVCFESPTAAILLPCRHPLCK
 SCSLASECPICRTKISDRLFAFSPS"

BASE COUNT

1160 a

740 c

842 g

950 t

ORIGIN

Query Match

Best Local Similarity

Matches 523; Conservative

1.8%; Score 176.4; DB 8;

Length 3692;

Pred. No. 5.7e-24;

Mismatches 396; Indels 39; Gaps 7;

Query

273

cttcaattcgatcggtatattcaattcgaatcaaacgaattcacaagaaga 332

Db

713

CTTACCCATTTTGAATTAAGTTTGGACACAGCACTACTATTGACGTTTATGATGATG 772

QY

333

tagcaatcactacatcagatcgcttcgagagatataatgcaaaatattgcaaa 392

Db

773

CTGCGAGACCAAGTAGTCAAGCGCAGCAATGGAAGCTTTAAGGAGCCGTTTGTCTTACG 832

QY

393

gacagacatcttcagaaagagctacacaaatgaggaacacaaatctcagtcata 452

Db

833

GTCTTACTAGCATGTGGAAGACACATACATGATGATGATGATGATGATGATGATGATG 892

QY

453

taccacagccatcacagaagttttaaattatcagaagatcacgaagaagaattc 512

Db

893

TACCGCTACCAATTAAGAGAGCTGTCATGATCCAGGATCTCCAGAGGAGGATTC 952

QY

513

tcttaagattctctatagaggaattacaaatgaaactgtgaagaagcctactgtgag 572

Db

953

TGCTCGGTTCATATCTTGAATATACATGAGGTGATGAATGATTTATCTG-----G 1006

QY

573

acagaagaagaagcccttggaattcgcgaggaattcaatagaagaagtgatgtcgtg 632

Db

1007

ATCCAAAGCCCAAAACTTACGCTGTAGAGAGGATCCAGG---CACTTATGTTGAGG 1063

QY

633

acctgactgaagaactgttaattgtctcgtgacatglaaatacagtgatcacaagaaggtg 692

Db

1064

GTATCAAGAGAGAGAGGTTGTTTGTCTGCGCATCAGATCATTCATTTGACGCTGGGG 1123

QY

693

aaaaaaacagacataatgagagacataaaatgaaatgaaatgaaatgaaatgaaatgaa 752

Db

1124

AAGACATGCTCATGTTGGTTCAATATTTTCAATTTGTTGACACACAGACATCAACATCA 1183

QY

753

tattagaatgattgttgaagaagccagagacagaatgatccacaatctcagaagaactgtg 812

Db

1184

TATTTACACTGATGCTTGAAGATG-----CTACTGGAGATGATGATG 1228

QY

813

atgagagctgcatgtaactcactgtaattgtgtagatctgtcgtcagtgaaagaaga 872

Db

1229

ATGGA---GTATCTTTCTCAACTCAATTTGATGACTTACCTGATCTGACAGCTTCCA 1285

QY

873

gocaaactgagcttgagagtgtagagcttaagaagaagccgaacacaaacccagagctgtc 932

Db

1286

---AAACTGAACAACTGATTTGAGAGAGAGGCTTCATACATCAACAAAGAGCTTTC 1342

QY

933

ttaacttgacaggtatataagaagcttagcagagccagagctgtgtgattatataaact 992

Db

1343

TAACTGTGGAAGTGTGATTTGAAAACTTAGTGAGGTTAAGGCACTACATTTCCA---T 1399

LOCUS	AY039966	3599 bp	mRNA	linear	PLN 24-JUN-2000
DEFINITION	Arabidopsis thaliana putative kinesin heavy chain (F3K23.14/At2g21380) mRNA, complete cds.				
ACCESSION	AY039966				
VERSION	AY039966.1				
KEYWORDS	FLI.CDNA.				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
AUTHORS	1 (bases 1 to 3599)				
REFERENCE	2 (bases 1 to 3599)				
AUTHORS	Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K., Dale,J.M., Gibson,H.A., Goldsmith,A.D., Jiang,P.X., Lee,J.M., Quach,H.L., Tang,C.C., Toriumi,M., Yu,G., Bowser,L., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Koeseana,E., Lam,B., Lin,J., Meyers,M.C., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinozaki,K., Shimizu,A., Shinzaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.				
TITLE	Full Length cDNA of gene F3K23.14/At2g21380 (GI:4567271)				
JOURNAL	Unpublished				
COMMENT	Direct Submission				
TITLE	Submitted (11-JUN-2001) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA				
JOURNAL	RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL.CDNA: 'RIKEN Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinozaki,K.				
FEATURES	The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K., Dale,J.M., Gibson,H.A., Goldsmith,A.D., Jiang,P.X., Lee,J.M., Quach,H.L., Tang,C.C., Toriumi,M., Yu,G., Bowser,L., Chen,H., Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C., Koeseana,E., Lam,B., Lin,J., Miranda,M., Nguyen,M., Palm,C.J., Shinozaki,K., Shimizu,A., Tracy,S.E., Davis,R.W., Ecker,J.R. and Theologis,A.				
SOURCE	Yamada,K. (SSP/PGEC) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGEC) contributed equally to this work as PIs.				
FEATURES	Location/Qualifiers				
gene	1. 3599				
5'UTR	/organism="Arabidopsis thaliana"				
CDS	/db_xref="taxon:3702"				
	/chromosome="2"				
	/clone="RAFL05-09-N13 (R10023)"				
	/note="This clone is in a modified p Bluescript vector (lambda ZAP) as a XhoI/SstI insert.				
	ecotype: Columbia				
	1. 3599				
	/gene="F3K23.14/At2g21380"				
	1. 119				
	/gene="F3K23.14/At2g21380"				
	120. 3296				
	/gene="F3K23.14/At2g21380"				
	/codon_start=1				
	/evidence=experimental				
	/product="putative kinesin heavy chain"				
	/protein_id="AAK6143.1"				
	/db_xref="GI:14532684"				
	/translation="MASSSSSTRGRSPSPSHRRPPSPVSSASTSSSLINNLPRSSSS"				

[illegible]

Db 1121 -CAAACTGAGACACGAGCTGAGAGAGAGGGTGCATACATCAACAGAGCTTCC 1179
 Qy 933 ttacccttgagacaggttataagaagccttagcagcgccagcgtggtgattataaact 992
 Db 1180 TGACCTTGGAACTGTGATTGGAAAACCTGACCTGAGGCAAGCAACT---CATGTCCCAT 1236
 Qy 993 acagagacagcaaacctaccagaatctccaaaatcattggaggaatgctaaacg 1052
 Db 1237 TTGCGGACTCTAAGCTAACTCGTCTTCGCAATCTCAGCTGAGTGCGCATGTAT 1296
 Qy 1053 ttataattgcaaatlaagccagttc-----tttgatagagctcctaagtaactc 1106
 Db 1297 CCCTCATATGATGATGATCTCTGCTGCTCAAGCATACCGAGAAACTCATATCTTGA 1356
 Qy 1107 agttgccagtaactgcaacatgtagaatactccctcctgtaatgagtcctgagc 1166
 Db 1357 AGTTTCCGACGAGGCGCAAGAGATTGAATTATGCTCAGGCATTAAGATCATGATG 1416
 Qy 1167 atgaagcgtgtgctaaagagctacagaaagaaatcttgatttaagaacaatagaa 1226
 Db 1417 ACAAGCATGATGATTAAGAAATATCAGAAAGAAATCTCGAAGCTCAAGTCAACTGATC 1476
 Qy 1227 attta 1231
 Db 1477 ACCTA 1481

RESULT 10

AF372930 1723 bp mRNA linear PLN 23-MAY-2001
 LOCUS Arabidopsis thaliana AT5g66310/K1L20_9 mRNA, partial cds.
 DEFINITION AF372930
 ACCESSION AF372930.1 GI:13937152
 VERSION FLI.CDNA.
 KEYWORDS
 SOURCE
 ORGANISM
 thale cress.
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 1723)

REFERENCE

AUTHORS Shinn, P., Chen, H., Cheuk, R., Kim, C.J., Meyers, M.C., Banh, J.,
 Bowser, L., Carninci, P., Chung, M.K., Goldsmith, A.D., Hayashizaki, Y.,
 Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawal, J.,
 Lam, B., Lee, J.M., Lin, J., Lin, S.X., Miranda, M., Narusaka, M.,
 Nguyen, M., Palm, C.J., Pham, P.K., Quach, H.L., Sakano, H., Sakurai, T.,
 Saitou, M., Seki, M., Southwick, A., Toriumi, M., Yamada, K., Yu, G.,
 Shinohaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.
 Arabidopsis cDNA clones
 Unpublished

TITLE

JOURNAL

AUTHORS

2 (bases 1 to 1723)

TITLE

JOURNAL

COMMENT

Submitted (20-APR-2001) Salk Institute Genomic Analysis Laboratory
 (SIGAL), Plant Biology Laboratory, The Salk Institute for
 Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
 USA
 RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and clustering of RAFI cDNAs (RAFI.CDNA: 'RIKEN
 Arabidopsis Full-Length cDNA'): Seki, M., Narusaka, M., Ishida, J.,
 Saitou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawal, J.,
 Hayashizaki, Y. and Shinohaki, K.

The Salk, Stanford, PCBC (SSP) Consortium members carried out the
 sequencing and annotation of the RAFI cDNAs: Shinn, P., Chen, H.,
 Cheuk, R., Kim, C.J., Koesema, E., Meyers, M.C., Tracy, S.E., Banh, J.,
 Bowser, L., Chung, M.K., Goldsmith, A.D., Jones, T., Karlin-Neumann, G.,
 Lam, B., Lee, J.M., Lin, J., Lin, S.X., Miranda, M., Nguyen, M.,

Palm, C.J., Pham, P.K., Quach, H.L., Sakano, H., Southwick, A.,
 Tang, C.C., Toriumi, M., Yamada, K., Yu, G., Davis, R.W., Theologis, A.,
 and Ecker, J.R.

Shinn, P. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to
 this work. Shinohaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk)
 contributed equally to this work as PIs.

FEATURES

source

1..1723

/organism="Arabidopsis thaliana"

/db_xref="taxon:3702"

/chromosome="5"

/clone="RAFI09-14-N03(RJ2318)"

/note="ecotype: Columbia"

48..>1723

/note="kinesin heavy chain DNA binding protein-like"

/codon_start=1

/product="AT5g66310/K1L20_9"

/protein_id="AAK50070.1"

/db_xref="GI:13937153"

/translation="MTTEDDDDMMTGGPGSTGKIVSVMPRLNDEKERNVPMWCT
 NNTTITRSHLSISERSMPSATTFDFRFSPECTROYIEGAGAEVAFSVYSNABV
 FAYQTSIGKTYTMSGITDCALVDVIGTIDHKKRERFLKRSAMEIYNESRDLISD
 TSPRLDDPEKGTVEKTEBELRDWHKFKELLSVCAQRIEETALNEYSRSHQI
 LRLVESIAREFSINDKFSLTATVNEIDLAGESEASQSISAGTRLEGGHINSLT
 LGTVIRKLSREKGTIIPROSKLTRILOSISGNARPAIICTNSPARIHQESNTLL
 FASCAKEVTTAAOVNVVMSDKALVKHILRELAKLESLRSQSISYSIDTALLTERD
 LEVERKKEVQLAQOLEQARSSELKDLRWVEE"

BASE COUNT

535 a 330 c 406 g 452 t

ORIGIN

Query Match 1.7%: Score 161.2; DB 8; Length 1723;
 Best Local Similarity 53.1%: Pred. No. 5.5e-21;

Matches 486; Conservative 0; Mismatches 388; Indels 42; Gaps 5;

Qy 273 cttaaatctgcatcggttatttaattcaccgaatacacaagtaaatcttaccagaaga 332
 Db 644 CTTACACATTTGACAGAGAGTGTTACCCGTGATGTTGACAGGAGGTATGAAACAG 703
 Qy 333 tagcagtcctcatcacacagtcagcttgcagagatataatgagcaaatattgcatacg 392
 Db 704 GGGCAAGAGAGTCCCTTCTCTGTTGACAGGCGGTAAATGCTAATGCTAATGCGATATG 763
 Qy 393 gacagacatcttcagcagaagcagacagatgagtagagcaccgaatcattgagcataa 452
 Db 764 GCCAACAACAGCAGTGGAAGACGTACACTATGAGCGGAATTAAGTACTGCTCC----- 813
 Qy 453 taaccacagcatcacaggaagttttaaatattatcagagagataccgacagagagtttc 512
 Db 814 -----GCCTTGGTTGATATATATGCTCATTTGACAGACAGCAAGAAAGAGAAATTC 865
 Qy 513 ttctaaggttcttataatgagagattcaatgaaactgtaagaactactggtgagtg 572
 Db 866 TTCTTAAGTTCTCTGCTATGAGATCTCAATGATGCTGTGAAGATCTCTTAAGTATACAG 925
 Qy 573 acagagaagaagaagcccttggaatctgcagagatatttaataagaacgtgtaigtctg 632
 Db 926 ATACTAG-----TCCACTCAGACTTTTATGATGACCCCTGAGAAAGGACAGTGTTCGA 979
 Qy 633 aactgactgaagaactgtaatggttcctgacacatgtaatacagtgagatcaaaaaaggtg 692
 Db 980 AACCTCAGAGAGAACTCGCGAGACTGGAATTCATTTTCAGAGAGTCTATCTGCTGCA 1039
 Qy 693 aaaaaaagaagactatgagagactaaatgaatgatacagtagtgcgttccacatacaa 752
 Db 1040 AAGCAAAAGCAAAATTTGTAGACAGCTCTTAATGAAGTATGAGTCCCGGCTCATCA 1099
 Qy 753 tatttagatgattgtttaaagcagacagaagaatgacccacaaatccagagagactg 812
 Db 1100 TTTTACAGATTACAGTAAAGCATAGCTGCTGATTTTCTACCAATGATATAGTTTACTA 1159
 Qy 813 atgagagctgcatglatctcacttgaaatttgtagatctctgctgcagtgaaagagcaa 872


```

/gene="SPKR85"
/codon_start=1
/product="SPKINESIN-II (KRP85/95) 85kd subunit"
/protein_id="AA16098.1"
/db_xref="GI:295246"

/translation="MPGSSGNDNVRVYVHCRPLNKEETOQGRKSVKMDENKCTVOV
TNPVAPSGEPKSTFEDTVFAPGAKQTDVYNQFAPLVDI1EGYNTTIPAYQQTG
KPTMEVRSQPELGI1IPNSFAHIEGHIAKEQENVEFLVRSYLEIENEVRLDGK
DOOHRELEVERPDLVYKDLISAFVNNADMDRIMTLGKNRSVGTANNENSSRSH
AIFPTLEERSMDGLDKEOHYRVGLHMDVLAGSROTKGTGORTGOREAKTKINLSIT
LVNYSLSVDGKSTHPIYRNSKTLRLIDLSGNKATVYMGANIGAPENYKDEITSLR
YANRAKIRKAKINEDPKALLREPKELBELKQISEGEGLDDEDSGSESGDE
EABGCVKRRKRNKRPKLSPELMAAMOK1DEBKALEKDMVEDRNTVHRELQ
RSESELHKAQDDK1LINEKLNALQKTLIVGVLDLASEQDL1DEQSALEMERAK
QESWRKMEEREQERMD1EKYS1LQDEAGTKTKLKVTMTLMQASEVADAQAEHQ
REMEALLENRELSRELRLSML1IDSF1PEQFEMIEQYVHNMDIDEMO1LCKCVAATG
NNMRKQTPVADKDKSLAYGEADLSNVFLTY1NLEGGGKMYPSQGSRRPTSSGRKPT
GKKKQASMSIDALLO"

BASE COUNT      722 a      455 c      625 g      413 t
ORIGIN

Query Match      1.6%; Score 153.2; DB 3; Length 2213;
Best Local Similarity 52.2%; Pred. No. 1.9e-19;
Matches 533; Conservative 0; Mismatches 448; Indels 41; Gaps 7;

OY 231 agcgtgaaacacacacattcccaagltgatyggacaaglttcaattcgatcg 290
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 176 AGGTACAAACCCCAATGACACCCAGCGGGAGCCCAAGTCTTCACCTTACACAG 235
OY 291 tatttaattccacgaatcaacagtcgaattaccagaagaatgacgactcatc 350
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 236 TGTTCGCGCCGGTCCAGACGACTATGTCTAACCAACACGCGCAGACCATCTG 295
OY 351 gatcagcttcgaggaataatgacaatattgcatatgacacatcttcagcga 410
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 296 ATGCTATCATTTGAGGGGTACACAGGACTATATTGCTACGCTACACACAGTAC 355
OY 411 agacgtacacaatgatyga-----acaccaattcattgagcataataccca 461
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 356 AGAGCTTACACCATGAGGATGACATCTCAACGAGAACTCAGGGGATCATCTCCA 415
OY 462 ccatacaggaagttt---aaatatccagaggaataccgaagaagagttcttca 518
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 416 CATTGCTCACATTTTGGACACACTTAAGAGCAAGAAATGTCAGGTTCCCTACT 475
OY 519 gagttcttataatgagatttacaatgaacctggaagacctgtygtatgacgaa 578
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 476 GCGTGTATATTTGGACATCTACAAATGAGAGGATTTATTAGCCAAAGAC---C 532
OY 579 gaaagaagcccttggaattcgcgagatttataagaacgtygtatgacctga 638
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 533 AGCGACGATAGATTAGAGGTCAAAAGCGCGGATGTTGGTGTCTATGTCAAGAC 592
OY 639 ctgaagaacttgtaatggttcccggaacatgtaataacgtygtatcaaaaag 698
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 593 CTGATTGCTGTGTAATTAACGACAGATGATGACAGAGTAATGACTCTAGGAAAC 652
OY 699 acagacatltgagagactaaatgatatcatagtagtcttcaatacatatttta 758
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 653 ACAGATGTGGGGGCGACAAACATGAAATCAATCAAGTTCTCGTTCCCAATTTT 712
OY 759 gaatgattgttgaagccggaacagaatgataccacaataatcagaagaactgt 818
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 713 CCATTAGCTTGGAAAGGAGTGACATGGGT-----CTGGATTAAGAACAGC 757
OY 819 ctgtcataggtatctcaatttgtagatcttctgacagtgaaagaagcca 878
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 758 ATGTCAAGGCGGAGGAGCTCACAATGTTGATCTTGGAGGTTCCGAGAGACACCA 817
OY 879 ctggaactggaagtgtagaagcgtgaagaagcgtcaaacacgacgttattat 938
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 818 CCGGTGCGACCGGACAAACGCTCAAGGAGGCGCACCAAGATCACTGTCTCTAC 877

```

```

OY 939 ttggacgattattagaagctttagcagcgccagcgctggtgattataactacag 998
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 878 TGGGTACGTCATCTCATCTCTGTTGATGATGTTAGACACTAC---ATCCCTACG 934
OY 999 acagcaactcaccagaattctccaaatcattcgtgaggaanaatgctaaacggtata 1058
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 935 ACTCCAGCTCACCACGAGCTCTTACAGACGACCTCTTGAGAGCAACGCCAAGTCA 994
OY 1059 ttgcaaatatcgccagt-----ttcctttagatgagactcgaagtaacctca 1112
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 995 GTGCCAATCTTGACCGACGACGAGTATACATGATGAGAACATTTAGCCACTAGAT 1054
OY 1113 ccagtactgcacaacatgtagaataatactcccatgtaatgagtgcttgatgata 1172
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1055 CAAMACGACGCAAGAACATCAAAAATTAAGCAAAAGATCATGAGATCCCAAGATG 1114
OY 1173 cgttgctaaaaagtgtagaagaagaatcttgatttaagaagaacttagagaatt 1232
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1115 TCTTG--AGAGAGTTCCAGAAAGAAATTGAAGATGAGAACAGATCTCAGATCT 1172
OY 1233 ag 1234
    ||
Db 1173 AG 1174

RESULT 14
AB072777      2759 bp      mRNA      linear      PRI 11-OCT-2001
LOCUS
DEFINITION
Macaca fascicularis testis cDNA clone:0tsa-20515, full insert
sequence.
ACCESSION
AB072777
VERSION
AB072777.1 GI:16041157
KEYWORDS
oligo capping; fts (full insert sequence).
SOURCE
Macaca fascicularis adult male testis cDNA to mRNA.
clone_lib:macaque testis cDNA library Qtsa clone:Qtsa-20515.
ORGANISM
Macaca fascicularis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
Cercopithecinae; Macaca.
1 (bases 1 to 2759)
AUTHORS
Hashimoto,K., Osada,N., Hida,M., Kusuda,J., Tanuma,R., Hirai,M.,
Tero,K. and Sugano,S.
Isolation of novel full-length cDNA clones from macaque testis cDNA
libraries
Unpublished
JOURNAL
REFERENCE
2 (bases 1 to 2759)
AUTHORS
Hashimoto,K., Osada,N., Hida,M., Kusuda,J. and Sugano,S.
DIRECT SUBMISSION
Submitted (09-OCT-2001) Katsuyuki Hashimoto, National Institute of
Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
(E-mail: khashi@nih.go.jp, URL: http://www.nih.go.jp/yoken/genbank/,
Tel:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181)
Lab host:
TOP10
Vector:
pME18S-FL3 (Acc.No. AB009864)
R. Site1:
DraIII (CACTGTCG)
R. Site2:
DraIII (CACCATGTC)
Description: 1st strand cDNA was primed with an oligo(dT) primer
[ATGTGGCTTTTCTTTTCTTTTCTTTT]; double-stranded cDNA was synthesized
using specific 5' and 3' primers and amplified by PCR. The PCR
product was digested with SfiI and size selection PCR product was performed to
exclude fragments <1.5kb. The SfiI-digested PCR product was cloned
into distinct DraIII sites of pME18S-FL3. XhoI sites just outside
the DraIII sites can be used to isolate the cDNA insert. Libraries
were constructed by oligo-capping method
(Sugano et al., University of Tokyo, Institute of Medical Science).
Custom primer used for sequencing
(5' end primer [CTGTGCTCAAAAGCGCG];
3' end primer [GACCTGCTCAGGACACA] ).
FEATURES
source
1..2759
/organism="Macaca fascicularis"

```



```

/translation="MSKSSSESVRVVHCRPMNSKELAGYERVVYDVKLGOVSVK
VKGATNELSKTFTTATYDYSNKOYEIVEMPRPVLVILGINTIRATGOTGTX
TYMEVGRDPEKRGVTPNSFEHITFHISRQNOULVRASTYEIETOELRDLSDO
SKRLERKRPDGVVYKDLSSFEVTSKVELEHVMNMGONRSYAGKMMNDSRSRAI
FMITIECSQIGLDGENHIRYKLNLDVLAGSEKOTGTGAGEELKATKATINLSLALG
NVTSAVDGSRSTHPIYRDSKLTIRLQVDSLGNAKTYMVAINTGASVNEVETLTLYS
NRKKNTKNRPVNEDEKDALREFOEIARLAKOLDRGVKRRRGRTAGGEVDE
DEDEBDLYEGDDVLTKDKEDYREOERIEIEKKALLEDSVAREKILKKEKREK
DLKREAMEMLNSKYKAMESKLINGKRYVHTNEOOKLLEOKRHEIARAEOKRERE
MOOQESODEENLEKETSLSLOEVDIKTKLKLFLSKQAVKAEIHDLOEPIKER
QELQONELTRELKHLIENFIPLEENKMNRCFEDEEDOKLHAIISLDNQ
MMRRPVSAIGYNRLCOHAKMSMMVRPDSYRAENIVYLEMDPSRTTRDYESPALAP
KVQAALEALQDDDDIQVDASTFDSATKRTKRPKSSRQPGANGSSAGSLIPQSRG
LVPK"

```

```

BASE COUNT      1111 a      866 c      884 g      900 t
ORIGIN
Query Match      1.5%; Score 148; DB 5; Length 3761;
Best Local Similarity 51.7%; Pred. No. 1.8e-18;
Matches 483; Conservative 0; Mismatches 415; Indels 36; Gaps 5;

```

```

QY 260 gatggacaagaagcttcaattcgatcgtglatltaattctcaagaatcaagaatcaa 319
DB 367 GAGGTGTCGCAAAACATTCATTTGTAGCGACTTATGACTCAACTCAAGCAAGTGGAG 426
QY 320 attaccaagaataagcagtcctatcatalcagatcagcttgcaggatataatggaca 379
DB 427 CTYATATGTGAACCTTTCAGGCCACTGGTGATTCGTTTGGCTTGAGACGGGACC 486
QY 380 atattgcatagcagacacatcttcaagaagcagctacacaaatgatggaca----- 433
DB 487 ATATTGCTACAGGTCACACAGAACAGAAAGACCTACACATGAGAGGAGTGGTGT 546
QY 434 ---ccaattcattggcataataccccaagccatacaggaaagttttaaatattcag 490
DB 547 GATCCAGAAAAGAGAGGGGTATACCCATTCATTGAGCACATATTACTCACATCTCA 606
QY 491 gagataccgaagaagatcttcttaagaagttcttataggagatttacaatgaact 550
DB 607 AGGTCTCAGAACGACGACTGCTGACAGGCTTCTACCTAGAGATTATCCAGSAGNA 666
QY 551 gtgaagaacctactgtgtagtagcagaagaagcccttgaagaattcgcagagattt 610
DB 667 ATCAAGAGACTTTTATCTAAGATCAGAG---CAAAAGCGCTGAGCTGAAGAGAGACT 723
QY 611 aatagaacatgtagtctgtagcagctgtagcagaactgttaagtgtcttcgtaacatgta 670
DB 724 GATATCGGGGTGTATGTCAAGGATTTGCTTCTTGTACCAAGAGGTGTGAAGAGANT 783
QY 671 atacagtgatcaaaaaagggtgtaaaaaacagacattatgagagactaaatgaatgat 730
DB 784 GAACATGTGATGAATGTAGGAAATCAAAACCGATCTGTTGGAGCGACCAATATGAATGA 843
QY 731 catagtagtcgtcacatacaatattagaaatgtagtltgaaagccgagacagaatgat 790
DB 844 CACAGGTCTCGCTCACACAGCCATCTTATGATCACAAATGAGTGTAGTCAGATTGTTTA 903
QY 791 cccacaatcagaagaactgtgtagagctgtcaatgtagtctcacttgaattgttagat 850
DB 904 GATGCGAANAACCCATTCGTTGTGA-----AAGCTGAACCTTGTTGAGAC 948
QY 851 ctgtcgtggaagtgaagaagcgaagccaaactgagctgaagtgtagtgaactlaagaagagc 910
DB 949 TTGGCAGGAGGTGAAGGCAAAAGACTGGAGCCCAAGGAGAAAGCGCTGAAGAGAGCA 1008
QY 911 tgaacatcaaccgacagctgttattccttggacaggttattaagaagcttagcgagcgc 970
DB 1009 ACTAAGATTAAATCTCTCTTCCGCACTAGGAATGTATCTCTGCTTGGTTGATGG- 1067
QY 971 caggctgtgtgatttataactacagacagaaactcacagaatttccaataattca 1030
DB 1068 --GAGAAGCACTCACATCCATACAGGATTCACAACTCACAGGTTACTACAGGACTCA 1125

```

```

QY 1031 ttggaggaatgctaaaaagttataatttgacaaattacgcagttc-----tttt 1084
DB 1126 CTGGGAGGAATGCCAAGACTGTCAATGTTGCCAATATAGGCCAGCTCTCAATATGTG 1185
QY 1085 gatgagacttcaagtaactcagtttgcagtaactgcgaagaatgtggaataactccc 1144
DB 1186 GAGGAGACATTTGACAACTCTCCGCTACTCTCAACAGAGCCAAAATATCAAGAACAAACCA 1245
QY 1145 catgttaatgaggtccctgagatgtagaagcgttgc 1178
DB 1246 AGACTGAATTAAGACCCCAAGATGCTTTATTCG 1279

```

Search completed: October 7, 2002, 14:08:34
Job time: 18586 sec

Tue Oct 8 10:22:05 2002

us-09-150-867-2.rge

Page 21

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 7, 2002, 09:33:18 ; Search time 158 Seconds
(without alignments)
14964.988 Million cell updates/sec

Title: US-09-150-867-2

Perfect score: 9626
Sequence: 1 gaatccgcgagtcgcatag.....ttaaaaaaaaaaacggaattc 9626

Scoring table: OLIGO.NUC
Gapop 60.0 , Gapext 60.0

Searched: 38353 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata/2/lna/5A_COMB.seq: *
2: /cgn2_6/prodata/2/lna/5B_COMB.seq: *
3: /cgn2_6/prodata/2/lna/5A_COMB.seq: *
4: /cgn2_6/prodata/2/lna/5B_COMB.seq: *
5: /cgn2_6/prodata/2/lna/5A_COMB.seq: *
6: /cgn2_6/prodata/2/lna/5B_COMB.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	0.2	10564	1 US-08-206-176-5	Sequence 5, Appl1
2	19	0.2	423	3 US-08-906-769-134	Sequence 134, App
3	19	0.2	423	3 US-08-906-616-134	Sequence 134, App
4	19	0.2	423	3 US-08-639-075A-134	Sequence 134, App
5	19	0.2	423	4 US-09-012-431-134	Sequence 134, App
6	19	0.2	423	4 US-09-012-692-134	Sequence 134, App
7	19	0.2	423	4 US-08-906-613-134	Sequence 134, App
8	19	0.2	445	1 US-08-207-904-14	Sequence 14, Appl
9	19	0.2	4080	4 US-09-328-111-665	Sequence 665, App
10	19	0.2	4080	1 US-08-375-300-1	Sequence 1, Appl1
11	19	0.2	4080	3 US-09-177-431-1	Sequence 1, Appl1
12	19	0.2	4080	5 PCT-US95-16930-1	Sequence 1, Appl1
13	19	0.2	31	2 US-08-859-998-170	Sequence 170, App
14	18	0.2	31	4 US-09-225-928-170	Sequence 170, App
15	18	0.2	315	2 US-08-743-200-1	Sequence 1, Appl1
16	18	0.2	432	2 US-09-001-944-11	Sequence 11, Appl1
17	18	0.2	432	2 US-09-240-004A-11	Sequence 11, Appl1
18	18	0.2	435	2 US-09-001-944-9	Sequence 9, Appl1
19	18	0.2	435	4 US-09-240-004A-9	Sequence 9, Appl1
20	18	0.2	450	4 US-09-240-004A-13	Sequence 13, Appl1
21	18	0.2	498	2 US-09-001-944-1	Sequence 1, Appl1
22	18	0.2	498	2 US-09-001-944-3	Sequence 3, Appl1
23	18	0.2	498	2 US-09-001-944-5	Sequence 5, Appl1
24	18	0.2	498	2 US-09-001-944-7	Sequence 7, Appl1
25	18	0.2	498	4 US-09-240-004A-1	Sequence 1, Appl1
26	18	0.2	498	4 US-09-240-004A-3	Sequence 3, Appl1
27	18	0.2	498	4 US-09-240-004A-5	Sequence 5, Appl1

28	18	0.2	498	4 US-09-240-004A-7	Sequence 7, Appl1
29	18	0.2	906	4 US-08-976-259-98	Sequence 98, Appl1
30	18	0.2	1558	1 US-08-455-550-7	Sequence 7, Appl1
31	18	0.2	1789	2 US-08-366-490-1	Sequence 1, Appl1
32	18	0.2	1789	3 US-08-860-483A-1	Sequence 1, Appl1
33	18	0.2	1923	1 US-08-077-939-14	Sequence 14, Appl1
34	18	0.2	1923	1 US-08-461-599-14	Sequence 14, Appl1
35	18	0.2	1923	1 US-08-461-621-14	Sequence 14, Appl1
36	18	0.2	1923	1 US-08-465-334-14	Sequence 14, Appl1
37	18	0.2	1926	1 US-07-901-703-12	Sequence 12, Appl1
38	18	0.2	1926	1 US-08-147-023-26	Sequence 26, Appl1
39	18	0.2	1926	1 US-08-278-729A-22	Sequence 22, Appl1
40	18	0.2	1926	1 US-08-480-528A-9	Sequence 9, Appl1
41	18	0.2	1926	1 US-08-479-666-9	Sequence 9, Appl1
42	18	0.2	1926	1 US-08-155-343A-22	Sequence 22, Appl1
43	18	0.2	1926	1 US-08-406-672-22	Sequence 22, Appl1
44	18	0.2	1926	1 US-08-643-563A-22	Sequence 22, Appl1
45	18	0.2	1926	1 US-08-447-570-26	Sequence 26, Appl1

ALIGNMENTS

RESULT 1
US-08-206-176-5
Sequence 5, Application US/08206176
Patent No. 563940
GENERAL INFORMATION:
APPLICANT: Garner, Ian
APPLICANT: Dalrymple, Michael A
APPLICANT: Prunkard, Donna E
APPLICANT: Foster, Donald C
TITLE OF INVENTION: Production of fibrinogen in Transgenic
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Zymogenetics, Inc.
STREET: 4225 Roosevelt Way, N.E.
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/206,176
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Parker, Gary E
REGISTRATION NUMBER: 31-648
REFERENCE/DOCKET NUMBER: 93-15
TELEPHONE: 206-547-8080 ext 322
TELEFAX: 206-548-2329
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 10564 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: human fibrinogen gamma chain
FEATURE:
NAME/KEY: CDS
LOCATION: join(1799..1876, 1973..2017, 2207..2390, 2510
..2603, 4211..4341, 4645..4778, 5758..5942, 7426
..7703, 9342..9571)
LOCATION: ..7703, 9342..9571)
US-08-206-176-5

Query Match 0.2%; Score 21; DB 1; Length 10564;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2975 TTCTCAAGCAGCTTGAGAA 2995
|||||
DB 3246 TTCTCAAGCAGCTTGAGAA 3266

RESULT 2

US-08-906-769-134
; Sequence 134, Application US/08906769
; Patent No. 6077687
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLUA PROTEASE PROTEINS, NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES AND USES THEREOF
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/906,769
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/639,075
; FILING DATE: 24-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 134:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..423
US-08-906-769-134

Query Match 0.2%; Score 19; DB 3; Length 423;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1824 AACTACAGTCAAGGAGA 1842
|||||
DB 222 AACTACAGTCAAGGAGA 240

RESULT 3

US-08-906-616-134
; Sequence 134, Application US/08906616
; Patent No. 6121035
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLUA AMINOPEPTIDASE PROTEINS AND USES THEREOF
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/906,616
; FILING DATE: 05-AUG-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 134:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..423
US-08-906-616-134

Query Match 0.2%; Score 19; DB 3; Length 423;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1824 AACTACAGTCAAGGAGA 1842
|||||
DB 222 AACTACAGTCAAGGAGA 240

RESULT 4

US-08-639-075A-134
; Sequence 134, Application US/08639075A
; Patent No. 6150125
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLUA PROTEASE PROTEINS, NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES AND USES THEREOF
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESSES:

ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/639,075A
FILING DATE: 24-APR-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 134:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..423
US-08-639-075A-134

Query Match 0.2%; Score 19; DB 3; Length 423;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1824 AACTACAGTCAAGAGAGA 1842
Db 222 AACTACAGTCAAGAGAGA 240

RESULT 5
US-09-012-431-134
Sequence 134, Application US/09012431
Patent No. 6180383
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
Rushlow, Keith E.
Wu Hunter, Shirley
Frank, Glenn R.
Stiegler, Gary
Gaines, Patrick J.
Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/012,431
FILING DATE: 23-Jan-1998

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/639,075
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 134:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..423
US-09-012-431-134

Query Match 0.2%; Score 19; DB 4; Length 423;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1824 AACTACAGTCAAGAGAGA 1842
Db 222 AACTACAGTCAAGAGAGA 240

RESULT 6
US-09-012-692-134
Sequence 134, Application US/09012692
Patent No. 6214579
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
Rushlow, Keith E.
Wu Hunter, Shirley
Frank, Glenn R.
Stiegler, Gary
Gaines, Patrick J.
Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/012,692
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/639,075
FILING DATE: 24-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700

TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 134:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..423
US-09-012-692-134

Query Match
Best Local Similarity 100.0%; Pred. No. 54;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1824 AACTACAGTCAAGAGGA 1842
|||||
Db 222 AACTACAGTCAAGAGGA 240

RESULT 7
US-08-906-613-134
Sequence 134, Application US/08906613
Patent No. 6232096
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Wu Hunter, Shirley
APPLICANT: Frank, Glenn R.
APPLICANT: Stiegler, Gary
APPLICANT: Gaines, Patrick J.
APPLICANT: Silver, Gary
TITLE OF INVENTION: FLUA PROTEASE PROTEINS, NUCLEIC ACID
TITLE OF INVENTION: MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/906,613
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/639,075
FILING DATE: 24-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 134:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..423

US-08-906-613-134

Query Match
Best Local Similarity 100.0%; Pred. No. 54;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1824 AACTACAGTCAAGAGGA 1842
|||||
Db 222 AACTACAGTCAAGAGGA 240

RESULT 8
US-08-207-904-14
Sequence 14, Application US/08207904
Patent No. 5477002
GENERAL INFORMATION:
APPLICANT: Tuttle, AnnMarie
APPLICANT: Crossland, Lyle D.
TITLE OF INVENTION: Antier-Specific cDNA Sequences, Genomic
TITLE OF INVENTION: DNA Sequences and Recombinant DNA Sequences
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/207,904
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/908,242
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Lazar, Steven R.
REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: CGC 1624
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8615
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 445 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Nicotiana tabacum
INDIVIDUAL ISOLATE: Ant68
FEATURE:
NAME/KEY: CDS
LOCATION: 2..445
US-08-207-904-14

Query Match
Best Local Similarity 100.0%; Pred. No. 54;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4924 CAATGTCAATTAAGTCAA 4942
|||||
Db 301 CAATGTCAATTAAGTCAA 319

RESULT 9
US-09-328-111-665
Sequence 665, Application US/09328111
Patent No. 6262333
GENERAL INFORMATION:
APPLICANT: Endege, Wilson O.
APPLICANT: Steinmann, Kathleen E.
APPLICANT: Astle, Jon H.
APPLICANT: Burgess, Christopher C.
APPLICANT: Bushnell, Steven E.
APPLICANT: Carroll III, Eddie
APPLICANT: Catino, Theodore J.
APPLICANT: Dertl, Adnan
APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Marcia E.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCD-257 (US)
CURRENT APPLICATION NUMBER: US/09/328,111
CURRENT FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: US 60/088,801
EARLIER FILING DATE: 1998-06-10
NUMBER OF SEQ ID NOS: 850
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 665
LENGTH: 658
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(658)
OTHER INFORMATION: n = A,T,C or G
US-09-328-111-665

Query Match 0.2%; Score 19; DB 4; Length 658;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3813 TGCAGCTGAGAACTCTGA 3831
|||||
DB 24 TGCAGCTGAGAACTCTGA 42

RESULT 10
US-08-375-300-1
Sequence 1, Application US/08375300
Patent No. 5679566
GENERAL INFORMATION:
APPLICANT: Feng, He
APPLICANT: Jacobson, Allan S.
TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE PRODUCTION IN
THE ABSENCE OF NONSENSE-MEDIATED MRNA DECAY FUNCTION
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street Suite 3100
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/375,300
FILING DATE: 20-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, J. P.
REGISTRATION NUMBER: 32,983

REFERENCE/DOCKET NUMBER: 04020/046001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)542-5070
TELEFAX: (617)542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4080 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-375-300-1

Query Match 0.2%; Score 19; DB 1; Length 4080;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4166 GAAAGAGTGAAGCTTCAA 4184
|||||
DB 862 GAAAGAGTGAAGCTTCAA 880

RESULT 11
US-09-177-431-1
Sequence 1, Application US/09177431
Patent No. 6071700
GENERAL INFORMATION:
APPLICANT: He, Feng
APPLICANT: Jacobson, Allan S.
TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE PRODUCTION IN THE
ABSENCE OF NONSENSE-MEDIATED MRNA DECAY FUNCTION
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/177,431
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/955,472
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, J. Peter
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 07917/050001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-9806
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4080 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-177-431-1

Query Match 0.2%; Score 19; DB 3; Length 4080;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4166 GAAAGAGTGAAGCTTCAA 4184

Db 862 GAAAGAGATGAGCTCAAA 880

RESULT 12

PCT-US95-16930-1

Sequence 1, Application PC/TUS9516930

GENERAL INFORMATION:

APPLICANT: UNIVERSITY OF MASSACHUSETTS MEDICAL SCHOOL

TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE

TITLE OF INVENTION: PRODUCTION IN THE ABSENCE OF

TITLE OF INVENTION: NONSENSE-MEDIATED mRNA DECA

TITLE OF INVENTION: FUNCTION

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson

STREET: 225 Franklin Street Suite 3100

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/16930

FILING DATE: 27-DEC-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/375,300

FILING DATE: 20-JAN-1995

ATTORNEY/AGENT INFORMATION:

NAME: Fasse, J Peter

REGISTRATION NUMBER: 32,983

REFERENCE/DOCKET NUMBER: 04020/046W01

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)542-5070

TELEFAX: (617)542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 4080 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

PCT-US95-16930-1

Query Match

Best Local Similarity 0.2%; Score 19; DB 5; Length 4080;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4166 GAAAGAGATGAGCTCAAA 4184

Db 862 GAAAGAGATGAGCTCAAA 880

RESULT 13

US-08-859-998-170

Sequence 170, Application US/08859998

Patent No. 5994076

GENERAL INFORMATION:

APPLICANT: Chenchik, Alex

APPLICANT: Bibilashvili, Robert

TITLE OF INVENTION: METHOD OF ASSAYING DIFFERENTIAL

TITLE OF INVENTION: EXPRESSION

NUMBER OF SEQUENCES: 1375

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson, P.C.

STREET: 2200 Sand Hill Road, Suite 100

CITY: Menlo Park

STATE: CA

COUNTRY: USA

ZIP: 94025

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows95

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/225,928

FILING DATE: 05-Jan-1999

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/859,998

FILING DATE: 21-MAY-1997

ATTORNEY/AGENT INFORMATION:

NAME: Fish & Richardson

REGISTRATION NUMBER: 09096/002001

REFERENCE/DOCKET NUMBER: 415-322-5070

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-854-0875

TELEFAX: 415-322-5070

INFORMATION FOR SEQ ID NO: 170:

SEQUENCE CHARACTERISTICS:

LENGTH: 31 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

OTHER INFORMATION: oligonucleotide primer

US-08-859-998-170

Query Match

Best Local Similarity 0.2%; Score 18; DB 2; Length 31;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5338 ATGCTGCTAATGAGAA 5355

Db 7 ATGCTGCTAATGAGAA 24

RESULT 14

US-09-225-928-170

Sequence 170, Application US/09225928

Patent No. 6352829

GENERAL INFORMATION:

APPLICANT: Chenchik, Alex

APPLICANT: Bibilashvili, Robert

TITLE OF INVENTION: METHOD OF ASSAYING DIFFERENTIAL

TITLE OF INVENTION: EXPRESSION

NUMBER OF SEQUENCES: 1375

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson, P.C.

STREET: 2200 Sand Hill Road, Suite 100

CITY: Menlo Park

STATE: CA

COUNTRY: USA

ZIP: 94025

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows95

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/225,928

FILING DATE: 05-Jan-1999

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/859,998

FILING DATE: 21-MAY-1997

ATTORNEY/AGENT INFORMATION:

NAME: Field, Bret E.
REGISTRATION NUMBER: 37,620
REFERENCE/DOCKET NUMBER: 09096/002001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-322-5070
TELEFAX: 415-854-0875
INFORMATION FOR SEQ ID NO: 170:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
OTHER INFORMATION: oligonucleotide primer
SEQUENCE DESCRIPTION: SEQ ID NO: 170:
US-09-225-928-170

Query Match 0.2%; Score 18; DB 4; Length 31;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5338 ATGCTGCTAATGAGAA 5355
|||||
Db 7 ATGCTGCTAATGAGAA 24

RESULT 15
US-08-743-200-1
Sequence 1, Application US/08743200
Patent No. 5861260
GENERAL INFORMATION:
APPLICANT: Doherty, Stephen J.
TITLE OF INVENTION: DIAGNOSTIC METHODS FOR SCREENING
TITLE OF INVENTION: PATENTS FOR SCLERODERMA
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/743,200
FILING DATE: 05-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, J. Peter
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 07917/025001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 315 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1...315
US-08-743-200-1

Query Match 0.2%; Score 18; DB 2; Length 315;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4408 ATTCAGTTAGAGAGGT 4425
|||||
Db 24 ATTCAGTTAGAGAGGT 41

Search completed: October 7, 2002, 11:14:00
Job time : 235 secs

THIS PAGE BLANK (USPTO)

FEATURES

Location/Qualifiers

1. 554

/organism="Xenopus laevis"

/db_xref="taxon:8355"

/clone="PBX0105E04"

/clone_lib="Blackshear/Soares normalized Xenopus egg library"

/sex="female"

/tissue_type="unfertilized egg"

/cell_type="unfertilized egg"

/dev_stage="unfertilized egg"

/lab_host="DH10B"

/note="Vector: pT73-Pac; Site 1: EcoRI; Site 2: NotI; PolyA-selected mRNA was prepared from unfertilized Xenopus laevis eggs. The library was constructed in the vector pT73-Pac as described in Bonaldo, M.F., Lennon, G. and Soares, M.B. 'Normalization and subtraction: two approaches to facilitate gene discovery', Genome Research 6:791-806, 1996. The first strand synthesis used a NotI-dt18 primer; double stranded cDNAs were ligated to EcoRI adapters, digested with NotI, and directionally cloned into the NotI and EcoRI-digested pT73-Pac vector. The library contained approximately 7.2 x 10⁵ recombinants, with average insert sizes of 1-1.5 kb."

BASE COUNT 177 a 107 c 129 g 141 t

Query Match

Best Local Similarity 100.0%; Pred. No. 2.2e-241; Length 554;

Matches 541; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 6662 GCTGTGATTCAGAGCAACACACTTGTCTCAGAGCTCTCTAGAGAGTTCAAAAGAA 7021
    |||||||
Db 14 GCTGTGATTCAGAGCAACACACTTGTCTCAGAGCTCTCTAGAGAGTTCAAAAGAA 73
QY 7022 ACTGAAGCACAAATATGATGTTAAATATTAAGAAATCTCTCATCCACGCTCC 7081
    |||||||
Db 74 ACTGAAGCACAAATATGATGTTAAATATTAAGAAATCTCTCATCCACGCTCC 133
QY 7082 AGATCTCTGGGACCTTGCAGCACTGATGTTAAATATTAAGAAATCTCTCATCCACGCTCC 133
    |||||||
Db 134 AGATCTCTGGGACCTTGCAGCACTGATGTTAAATATTAAGAAATCTCTCATCCACGCTCC 193
QY 7142 CTGAACAATTTAAGGTTGATATACGAACTGCTGAGTCAAAAGAGATCATAGCTTATC 7201
    |||||||
Db 194 CTGAACAATTTAAGGTTGATATACGAACTGCTGAGTCAAAAGAGATCATAGCTTATC 253
QY 7202 AAAGATTTAGAGAGAGCTTGTCTGAGCAAAAGAGGATGATGAGCTGCAGCTCAA 7261
    |||||||
Db 254 AAAGATTTAGAGAGAGCTTGTCTGAGCAAAAGAGGATGATGAGCTGCAGCTCAA 313
QY 7262 CTGAGTGTGGAGCAGCAGGCAAAATGTCGATTCCTGATCTGAGAACTCAAG 7321
    |||||||
Db 314 CTGAGTGTGGAGCAGCAGGCAAAATGTCGATTCCTGATCTGAGAACTCAAG 373
QY 7322 TTCTGTGAATTTGAATCTTGAATGATTTCTTTTAAAGCAAAATATTAATTCAGAGT 7381
    |||||||
Db 374 TTCTGTGAATTTGAATCTTGAATGATTTCTTTTAAAGCAAAATATTAATTCAGAGT 433
QY 7382 GTCAGAGATGACTTTTCAAGAGTGCAGATTTCTTAATCAAGTAGATCAACACTGCAA 7441
    |||||||
Db 434 GTCAGAGATGACTTTTCAAGAGTGCAGATTTCTTAATCAAGTAGATCAACACTGCAA 493
QY 7442 GAAGAGCTTGAGCAAGAAAGCTTTATGAGAGTGTGGAGAAATTTGGAGATCTGCAC 7501
    |||||||
Db 494 GAAGAGCTTGAGCAAGAAAGCTTTATGAGAGTGTGGAGAAATTTGGAGATCTGCAC 553
QY 7502 G 7502
    |
Db 554 G 554

```

RESULT 2

B1349745

LOCUS

DEFINITION

B1349745 426 bp mRNA linear EST 31-JUL-2001
 dae63c09.y3 Blackshear/Soares normalized Xenopus egg library
 Xenopus laevis cDNA clone IMAGE:4678672 5' similar to TR:042263
 O42263 KINESIN-RELATED PROTEIN. ; mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Xenopus laevis

African clawed frog.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Sandy Clifton, Ph.D.
 Washu Xenopus EST project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Library constructed by Bento Soares and M. Fatima Bonaldo
 (University of Iowa). DNA Sequencing by: Washington University
 Genome Sequencing Center
 Clone distribution: Xenopus clones from this library are available
 through the I.M.A.G.E. Consortium/LNL at: info@image.llnl.gov
 Putative full length read
 vector to vector length is 496.

FEATURES

source

Location/Qualifiers

1. 426

/organism="Xenopus laevis"

/db_xref="taxon:8355"

/clone="IMAGE:4678672"

/clone_lib="Blackshear/Soares normalized Xenopus egg library"

/sex="female"

/tissue_type="unfertilized egg"

/cell_type="unfertilized egg"

/dev_stage="unfertilized egg"

/lab_host="DH10B"

/note="Vector: pT73-Pac; Site 1: EcoRI; Site 2: NotI; PolyA-selected mRNA was prepared from unfertilized Xenopus laevis eggs. The library was constructed in the vector pT73-Pac as described in Bonaldo, M.F., Lennon, G. and Soares, M.B. 'Normalization and subtraction: two approaches to facilitate gene discovery', Genome Research 6:791-806, 1996. The first strand synthesis used a NotI-dt18 primer; double stranded cDNAs were ligated to EcoRI adapters, digested with NotI, and directionally cloned into the NotI and EcoRI-digested pT73-Pac vector. The library contained approximately 7.2 x 10⁵ recombinants, with average insert sizes of 1-1.5 kb."

BASE COUNT 145 a 73 c 105 g 103 t

ORIGIN

Query Match

Best Local Similarity 4.4%; Score 424; DB 10; Length 426;

Matches 424; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 7239 GGCATGATGAGTGCAGCTCCAACTGAGTGTGGAGCAGCAGCAGCAAAATGTCGG 7298
    |||||||
Db 1 GGCATGATGAGTGCAGCTCCAACTGAGTGTGGAGCAGCAGCAGCAAAATGTCGG 60
QY 7299 ATTCGATCTGAGAGAACTCAAGTTCTGTGAATTTGAATTTGAGTACTTTT 7358
    |||||||
Db 61 ATTCGATCTGAGAGAACTCAAGTTCTGTGAATTTGAATTTGAGTACTTTT 120

```

```

OY 7359 AAAAGCAATATTAATTCAGAGTGTCCAGATGACTTTTCAGAGGTGACGATATTCCTAA 7418
DB 121 AAAAGCAATATTAATTCAGAGTGTCCAGATGACTTTTCAGAGGTGACGATATTCCTAA 180
OY 7419 ATCAAGTAGATCAACACATGCAAGAAGAGCTTGAGCACAAGAAAGGCTTTATGCACTGCT 7478
DB 181 ATCAAGTAGATCAACACATGCAAGAAGAGCTTGAGCACAAGAAAGGCTTTATGCACTGCT 240
OY 7479 TGGAGGAATTTGGAGATCTGCACAGCTGCATGCTAAGAAACTGAGTGAAGGATGCAACAGG 7538
DB 241 TGGAGGAATTTGGAGATCTGCACAGCTGCATGCTAAGAAACTGAGTGAAGGATGCAACAGG 300
OY 7539 AAAATAGGCGCATGCTCTTACCATATACAGCTCTTAACAAAAGGCTAAAGGCGAGTGTTC 7598
DB 301 AAAATAGGCGCATGCTCTTACCATATACAGCTCTTAACAAAAGGCTAAAGGCGAGTGTTC 360
OY 7599 AGTCGAAAATACACAGTGAATACCGTATATCTGAACCGAGTTTGAAGGAAATTCGCAAG 7658
DB 361 AGTCGAAAATACACAGTGAATACCGTATATCTGAACCGAGTTTGAAGGAAATTCGCAAG 420
OY 7659 AGAA 7662
DB 421 AGAA 424

RESULT 3
BM262378 636 bp mRNA linear EST 18-DEC-2001
DEFINITION daq40c12.y3 Blackshear/Soares normalized Xenopus egg library
042263 KINESIN-RELATED PROTEIN. ; mRNA sequence.
ACCESSION BM262378
VERSION BM262378.1 GI:17925418
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 636)
AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
Waterston,R. and Wilson,R.
Washu Xenopus EST project, 1999
Unpublished (1999)
Other_ESTS: daq40c12.x3
Contact: Sandy Clifton, Ph.D.
Washu Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Library constructed by Bento Soares and M. Fatima Bonaldo
(University of Iowa). DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LIML at: info@image.liml.gov
Seq primer: -40RP from Gibco
High quality sequence stop: 493.
FEATURES
SOURCE
location/Qualifiers
1..636
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:4783871"
/library="
libray="
/sex="female"
/tissue_type="unfertilized egg"
/cell_type="unfertilized egg"
/dev_stage="unfertilized egg"
/lab_host="DH10B"

```

```

/note="Vector: pT7T3-Pac; Site.1: EcoRI; Site.2: NotI;
PolyA-selected mRNA was prepared from unfertilized Xenopus
laevis eggs. The library was constructed from unfertilized Xenopus
pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and
Soares, M.B. 'Normalization and subtraction: two
approaches to facilitate gene discovery', Genome Research
6:791-806, 1996. The first strand synthesis used a
NotI-dT18 primer: double stranded cDNAs were ligated to
EcoRI adapters, digested with NotI, and directionally
cloned into the NotI and EcoRI-digested pT7T3-Pac vector.
The library contained approximately 7.2 x 105
recombinants, with average insert sizes of 1-1.5 kb."
BASE COUNT 185 a 126 c 125 g 200 t
ORIGIN
Query Match 4.3%; Score 417; DB 10; Length 636;
Best Local Similarity 100.0%; Pred. No. 2.2e-183;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 8876 CAGTTTTCATATTCATAGCTTGCTGATTTTTCAGAGCTCAACACAGACGACCAAT 8935
DB 97 CAGTTTTCATATTCATAGCTTGCTGATTTTTCAGAGCTCAACACAGACGACCAAT 156
OY 8936 GACAAAGTCAGGCTGAGAACTGCTGATGAAGCAAAAAAGAAAGGCGACCTGAATG 8995
DB 157 GACAAAGTCAGGCTGAGAACTGCTGATGAAGCAAAAAAGAAAGGCGACCTGAATG 216
OY 8996 AAAACATCTAGATCCCTGATACATCTGCTCTCTGCAAGAGACTTGCTACTCT 9055
DB 217 AAAACATCTAGATCCCTGATACATCTGCTCTCTGCAAGAGACTTGCTACTCT 276
OY 9056 GCCTTTTGTAGGAAGAAACACATGACATGCTGCTGCAAGAGACTTGCTACTGCA 9115
DB 277 GCCTTTTGTAGGAAGAAACACATGACATGCTGCTGCAAGAGACTTGCTACTGCA 336
OY 9116 AGCAAAAGTTGTTCTTATAGTAATACATGCTGTTGGCGAGTGTGACGCTTAAATA 9175
DB 337 AGCAAAAGTTGTTCTTATAGTAATACATGCTGTTGGCGAGTGTGACGCTTAAATA 396
OY 9176 AAGTGCAATAGCTGTCAGCTGATTTTATATGTTATGCTGCTGATTTTACACTTTTA 9235
DB 397 AAGTGCAATAGCTGTCAGCTGATTTTATATGTTATGCTGCTGATTTTACACTTTTA 456
OY 9236 AGTCCCTTACTTCATATTTGGCTCATCTGTAATTTCTTTGTTGGCCACACACA 9292
DB 457 AGTCCCTTACTTCATATTTGGCTCATCTGTAATTTCTTTGTTGGCCACACACA 513

RESULT 4
BM262071/c 583 bp mRNA linear EST 18-DEC-2001
LOCUS daq40c12.x3 Blackshear/Soares normalized Xenopus egg library
DEFINITION Xenopus laevis cDNA clone IMAGE:4783871 3', mRNA sequence.
ACCESSION BM262071
VERSION BM262071.1 GI:17925111
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 583)
AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
Waterston,R. and Wilson,R.
Washu Xenopus EST project, 1999
Unpublished (1999)
Contact: Sandy Clifton, Ph.D.
Washu Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800

```

Fax: 314 286 1810
 Email: estewatson.wustl.edu
 Library constructed by Bento Soares and M. Fatima Bonaldo
 (University of Iowa). DNA Sequencing by: Washington University
 Genome Sequencing Center
 Clone distribution: Xenopus clones from this library are available
 through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
 Seq primer: -400P from GIBCO
 High quality sequence stop: 510.

FEATURES
 source
 1. 583
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /clone="IMAGE:4783871"
 /clone_lib="BlackShear/Soares normalized Xenopus egg
 library"
 /sex="female"
 /tissue_type="unfertilized egg"
 /cell_type="unfertilized egg"
 /dev_stage="unfertilized egg"
 /lab_host="DH10B"
 /note="Vector: pT73-Pac; Site_1: EcoRI; Site_2: NotI;
 Polya-selected mRNA was prepared from unfertilized Xenopus
 laevis eggs. The library was constructed in the vector
 pT73-Pac as described in Bonaldo, M.F., Lennon, G. and
 Soares, M.B. 'Normalization and subtraction: two
 approaches to facilitate gene discovery', Genome Research
 6:791-806, 1996. The first strand synthesis used a
 NotI-dn18 primer; double stranded cDNAs were ligated to
 EcoRI adapters, digested with NotI, and directionally
 cloned into the NotI and EcoRI-digested pT73-Pac vector.
 The library contained approximately 7.2 x 10⁵
 recombinants, with average insert sizes of 1-1.5 kb."

BASE COUNT
 201 a 103 c 99 g 180 t

ORIGIN

Query Match
 Best Local Similarity 100.0%; Score 228; DB 10; Length 583;
 Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9065 TAGGAAGAACACTAGAAAGTGCATGCTAAAGAGAGCTGCTGGAAGCAAAAGT 9124
 |||||||
 Db 583 TAGGAAGAACACTAGAAAGTGCATGCTAAAGAGAGCTGCTGGAAGCAAAAGT 9124
 Qy 9125 TGTCTTAACTAGTAATCACTGCTGCGAGAGTGTACGCTTAAATAAGTGCAT 9184
 |||||||
 Db 523 TGTCTTAACTAGTAATCACTGCTGCGAGAGTGTACGCTTAAATAAGTGCAT 9184
 Qy 9185 AGCTGCACGCTCTAATTTATATGCTGCTGTAATTTGTTACACTTTAAGTCCCTG 9244
 |||||||
 Db 463 AGCTGCACGCTCTAATTTATATGCTGCTGTAATTTGTTACACTTTAAGTCCCTG 9244
 Qy 9245 ACTTCATATTTGGCTCATCTGTAAGTTCTTTGTTGCGACACACACA 9292
 |||||||
 Db 403 ACTTCATATTTGGCTCATCTGTAAGTTCTTTGTTGCGACACACACA 356

RESULT 5
 BG408399 430 bp mRNA linear EST 13-MAR-2001
 LOCUS dd01b05.y1 NICHD XGC 001 Xenopus laevis cDNA clone IMAGE:3405249 5'
 DEFINITION similar to TR:042263 042263 KINESIN-RELATED PROTEIN. ; mRNA
 ACCESSION BG408399
 VERSION BG408399.1 GI:13314748
 KEYWORDS EST.
 SOURCE African clawed frog.
 ORGANISM Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 Xenopodinae; Xenopus.
 REFERENCE 1 (bases 1 to 430)
 AUTHORS NCI-GAP <http://www.ncbi.nlm.nih.gov/ncigap>.

TITLE
 JOURNAL
 COMMENT
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: c9aps-r@mail.nih.gov
 Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: Xenopus clones from this library are available
 through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
 Seq primer: -40RP from GIBCO
 High quality sequence stop: 216.

FEATURES
 source
 1. 430
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /clone="IMAGE:3405249"
 /clone_lib="NICHD XGC 001"
 /tissue_type="oocytes"
 /lab_host="DH10B (phage-resistant)"
 /note="Vector: PCMV-SPO6; Site_1: NotI; Site_2: SalI;
 Cloned unidirectionally. Primer: Oligo dt. Average insert
 size 2.2 kb. Constructed by Life Technologies."

BASE COUNT
 167 a 76 c 105 g 81 t

ORIGIN

Query Match
 Best Local Similarity 99.6%; Score 181; DB 10; Length 430;
 Matches 231; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 7912 AAACGCAATCAGCTGAACCTAACTAAAGCAAGATATCTCAGCGTCAATGAAGA 7971
 |||||||
 Db 1 AAACGCAATCAGCTGAACCTAACTAAAGCAAGATATCTCAGCGTCAATGAAGA 60
 Qy 7972 AATAGAAACCTACAGAAATGTTGCCAAGTGCTACCTATTAAGAAGAAATGA 8031
 |||||||
 Db 61 AATAGAAACCTACAGAAATGTTGCCAAGTGCTACCTATTAAGAAGAAATGA 120
 Qy 8032 CAACCTTAAACTAGAGTGTAAAGATGAATGAATGAATGAATGAATGAATGA 8091
 |||||||
 Db 121 CAACCTTAAACTAGAGTGTAAAGATGAATGAATGAATGAATGAATGAATGA 180
 Qy 8092 AGACCAAGAGATGCTCTAATAGCTTTGTTGCAAGATGAAGAGAGGC 8143
 |||||||
 Db 181 AGACCAAGAGATGCTCTAATAGCTTTGTTGCAAGATGAAGAGAGGC 232

RESULT 6
 BG371437/c 545 bp mRNA linear EST 09-MAR-2001
 LOCUS dac4b12.x1 RIKEN Xenopus egg Xenopus laevis cDNA clone
 DEFINITION IMAGE:4433519 3' similar to TR:042263 042263 KINESIN-RELATED
 ACCESSION BG371437
 VERSION BG371437.1 GI:13267974
 KEYWORDS EST.
 SOURCE African clawed frog.
 ORGANISM Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 Xenopodinae; Xenopus.
 REFERENCE 1 (bases 1 to 545)
 AUTHORS Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hallier, L., Pape, D.,
 Martin, J., Wylie, T., Underwood, K., Theisling, B., Bowers, Y., Person
 B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.,
 Waterston, R. and Wilson, R.
 Washu Xenopus EST project, 1999
 Unpublished (1999)
 Contact: Sandy Clifton, Ph.D.
 Washu Xenopus EST project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estewatson.wustl.edu
 Library constructed by Piero Carninci (RIKEN, Genome Science Laboratory, Japan: <http://genome.riken.go.jp/>) using egg RNA supplied by Marc Kirschner (Harvard Medical School, Dept. of Cell Biology) DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: Xenopus clones from this library are available through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
 Seq primer: -40UP from G1bco
 High quality sequence stop: 376.
 Location/Qualifiers

FEATURES

1..545
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /clone="IMAGE:443519"
 /clone_1lb="RIKEN Xenopus egg"
 /tissue_type="egg (unfertilized)"
 /lab_host="DH10B"
 /note="Vector: PCS2G (custom); Site_1: BamHI; Site_2: XhoI; CDNA oligo-dr primed using oligo 5'-GAGAGAGAGAGATCCATCTGAGAGATTTTCTTTTCTTAA-3' and directionally cloned into XhoI and BamHI sites using the 5' adaptor 5'-ACGAGAGAGCTCGAGCTCTATAGTGACACTAGACCA-3'. Size selection was done automatically by the cloning vector (lambda pLC-II, P. Carninci et al., unpublished). Library was amplified at the phage stage and normalized (ROT value = 5). This library was constructed using the CapTrapper method and Trehalose thermoactivated reverse transcriptase to enrich for full-length clones. Library constructed by Piero Carninci (RIKEN, Genome Science Laboratory, Japan: <http://genome.riken.go.jp/>) using egg RNA supplied by Marc Kirschner (Harvard Medical School, Dept. of Cell Biology). References: Methods Enzymol. 303:19-44 and Genome Res. 10:1617-1630 Please contact Marc Kirschner (marc@hms.harvard.edu) for information on obtaining aliquots of this cDNA library."

BASE COUNT 139 a 104 c 108 g 194 t
 ORIGIN

Query Match 1.9%; Score 179; DB 10; Length 545;
 Best Local Similarity 99.1%; Pred. No. 4.6e-72;
 Matches 329; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1508 GCAAGTCTCGACATGCTTCATTTCCAGAAATGATGACTCTTTGTACAGAGTTT 1567
 |||
 DB 338 GCAAGTCTCGACATGCTTCATTTCCAGAAATGATGACTCTTTGTACAGAGTTT 279
 QY 1568 TCTGATTTGATGACGCCCTCTCCATGATGACAGCAATGGAATAGATGACAGATGGAAT 1627
 |||
 DB 278 TTTGATTTTGTAGACGCCCTCTCCATGATGACAGCAATGGAATAGATGACAGATGGAAT 219
 QY 1628 TTAGCCAGTAAAGTAACTACAGAGAAAAGACATCACTTCATCAATCAATGATAGACTTT 1687
 |||
 DB 218 TTAGCCAGTAAAGTAACTACAGAGAAAAGACATCACTTCATCAATCAATGATAGACTTT 159
 QY 1688 GGACAGATTTTGACAGTGTTCAGTTTCATGATTTCTTAAGAAAAACAGACTACATAC 1747
 |||
 DB 158 GGACAGATTTTGACAGTGTTCAGTTTCATGATTTCTTAAGAAAAACAGACTACATAC 99
 QY 1748 CTCGCCAAGACTCTGGTATATGGCTGAATGAGAAAAGCTTTTGAAGAAAAGATC 1807
 |||
 DB 98 CTCGCCAAGACTCTGGTATATGGCTGAATGAGAAAAGCTTTTGAAGAAAAGATC 39
 QY 1808 ACAAGCTCCAGCAACACTACAGTCAAGA 1839
 |||
 DB 38 ACAAGCTCCAGCAACACTACAGTCAAGA 7

RESULT 7
 BG486505/c

LOCUS BG486505 408 bp mRNA linear EST 22-MAR-2001
 DEFINITION dd01b05.x1 NICHD XGC 001 Xenopus laevis cDNA clone IMAGE:3405249

ACCESSION BG486505
 VERSION BG486505.1 GI:13434082
 KEYWORDS EST.
 SOURCE African clawed frog.
 ORGANISM Xenopus laevis

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae; Xenopodinae; Xenopus.
 1 (bases 1 to 408)
 NCBI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 Tumor Gene Index

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 TITLE Unpublished (1997)
 JOURNAL Other-ESTs: dd01b05.y1
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Martha Reibert, Steven L. Klein, Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: Xenopus clones from this library are available through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
 Seq primer: -40UP from G1bco
 High quality sequence stop: 320.
 Location/Qualifiers

FEATURES

1..408
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /clone="IMAGE:3405249"
 /clone_1lb="NICHD XGC 001"
 /tissue_type="oocytes"
 /lab_host="DH10B (phage-resistant)"
 /note="Vector: PCMV-SPORT6; Site_1: NotI; Site_2: SalI; cloned unidirectionally. Primer: Oligo dr. Average insert size 2.2 kb. Constructed by Life Technologies."

BASE COUNT 149 a 62 c 70 g 127 t
 ORIGIN

Query Match 1.8%; Score 172; DB 10; Length 408;
 Best Local Similarity 100.0%; Pred. No. 9e-69;
 Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9290 ACACACTGCTGAAAGATGACATTTGCAGTGTATGTTTACTGACTGCTCTGGGCG 9349
 |||
 DB 334 ACACACTGCTGAAAGATGACATTTGCAGTGTATGTTTACTGACTGCTCTGGGCG 275
 QY 9350 CATCATGCAAAACACCATTAAGTGTGCAATGTTTCTACTTATATATGCTGAC 9409
 |||
 DB 274 CATCATGCAAAACACCATTAAGTGTGCAATGTTTCTACTTATATATGCTGAC 215
 QY 9410 TTTGTGAATAGAAATACTACAAAAGATAGGCAAAAGTTGTTTCTCACT 9461
 |||
 DB 214 TTTGTGAATAGAAATACTACAAAAGATAGGCAAAAGTTGTTTCTCACT 163

RESULT 8
 AM636063 457 bp mRNA linear EST 26-APR-2001
 LOCUS b142c09.w1 Blackshear/Soares normalized Xenopus egg library Xenopus
 DEFINITION laevis cDNA clone PBX0042C09 5', mRNA sequence.
 ACCESSION AM636063
 VERSION AM636063.1 GI:7393144
 KEYWORDS EST.
 SOURCE African clawed frog.
 ORGANISM Xenopus laevis

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae; Xenopodinae; Xenopus.
 1 (bases 1 to 457)
 Blackshear, P.J., Lal, W.S., Thorn, J.M., Kennington, E.A., Staffa, N.G.


```

/clone.lib="XGC-egg"
/dev.stage="egg"
/lab.host="Escherichia coli XL1-blue"
/notes="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from 5ug of poly A+ RNA from egg.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end"
BASE COUNT      169 a      97 c      116 g      139 t
ORIGIN

Query Match      0.5%; Score 45; DB 9; Length 521;
Best Local Similarity 100.0%; Pred. No. 2.2e-09;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1538 GAATATGATGACTCTGTTGTACAGAGTTTTCGATTGTGATGAC 1582
|||||
Db 108 GAATATGATGACTCTGTTGTACAGAGTTTTCGATTGTGATGAC 152

RESULT 11
AL659719      670 bp      mRNA      linear      EST 13-DEC-2001
LOCUS      AL659719 XGC-neurula silurana tropicalis cDNA clone TNeu047p06 5',
DEFINITION      mRNA sequence.
ACCESSION      AL659719.1 GI:17673545
VERSION      EST.
KEYWORDS      western clawed frog.
SOURCE      Silurana tropicalis
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Silurana.
1 (bases 1 to 670)
Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (10_2001)
Unpublished (2001)
Contact: Huckle E
Sanger Centre
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: TNeu047p06.sp6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
Location/Qualifiers
1..670
/organism="Silurana tropicalis"
/db_xref="taxon:8364"
/clone="TNeu047p06"
/clone.lib="XGC-neurula"
/dev.stage="neurula"
/lab.host="Escherichia coli DH10B"
/notes="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from 5ug of poly A+ RNA from neurula.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end."
BASE COUNT      218 a      158 c      135 g      159 t
ORIGIN

Query Match      0.4%; Score 40; DB 9; Length 670;
Best Local Similarity 100.0%; Pred. No. 4.6e-07;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8328 CACATTATAGAGAATATCATCTTTATTCACGACTAT 8367
|||||
Db 86 CACATTATAGAGAATATCATCTTTATTCACGACTAT 125

RESULT 12
AL632671      648 bp      mRNA      linear      EST 02-NOV-2001
LOCUS      AL632671 XGC-gastrula silurana tropicalis cDNA clone Tgas020f09 5',
DEFINITION      mRNA sequence.
ACCESSION      AL632671.1 GI:16602154
VERSION      EST.
KEYWORDS      western clawed frog.
SOURCE      Silurana tropicalis
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Silurana.
1 (bases 1 to 648)
Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (10_2001)
Unpublished (2001)
Contact: Huckle E
Sanger Centre
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: Tgas020f09.sp6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
Location/Qualifiers
1..648
/organism="Silurana tropicalis"
/db_xref="taxon:8364"
/clone="Tgas020f09"
/clone.lib="XGC-gastrula"
/dev.stage="gastrula (stages 10.5-13 mixed)"
/lab.host="Escherichia coli XL1-blue"
/notes="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from 5ug of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pCS107 with EcoRI at the 5' end and NotI at the 3'
end."
BASE COUNT      252 a      116 c      143 g      134 t      3 others
ORIGIN

Query Match      0.4%; Score 37; DB 9; Length 648;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4426 GTTCAGTGTTTCACAGAGCTGTAGATGATAGAG 4462
|||||
Db 595 GTTCAGTGTTTCACAGAGCTGTAGATGATAGAG 631

RESULT 13
AL638330      681 bp      mRNA      linear      EST 12-DEC-2001
LOCUS      AL638330 XGC-neurula silurana tropicalis cDNA clone TNeu018h15 5',
DEFINITION      mRNA sequence.
ACCESSION      AL638330.1 GI:16790309
VERSION      EST.
KEYWORDS      western clawed frog.
SOURCE      Silurana tropicalis
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Silurana.
1 (bases 1 to 681)
Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (10_2001)
Unpublished (2001)
Contact: Huckle E
Sanger Centre
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: TNeu018h15.sp6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
Location/Qualifiers
1..681
/organism="Silurana tropicalis"
/db_xref="taxon:8364"
/clone="TNeu018h15"
/clone.lib="XGC-neurula"
/dev.stage="neurula"
/lab.host="Escherichia coli XL1-blue"
/notes="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from 5ug of poly A+ RNA from neurula.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end."
BASE COUNT      218 a      158 c      135 g      159 t
ORIGIN

```

```

FEATURES
  source
    location/Qualifiers
    1..681
      /organism="Silurana tropicalis"
      /db_xref="taxon:8364"
      /clone="TNeu018h15"
      /clone_1id="XGC-neurula"
      /dev_stage="neurula"
      /lab_host="Escherichia coli DH10B"
      /note="Vector: PCS107; Site_1: ECORI; Site_2: NOTI; cDNA
      was oligo dt primed from 5ug of poly A+ RNA from neurula.
      ECORI-NOTI cut cDNA was then ligated into PCS107 with
      EcorI at the 5' end and NotI at the 3' end."
      EcorI at 129 c 158 g 169 t 1 others
BASE COUNT      224 a 129 c 158 g 169 t 1 others
ORIGIN
Query Match      0.4%; Score 36; DB 9; Length 681;
Best Local Similarity 100.0%; Pred. No. 3.4e-05;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 284 GATCGTGTATTATTCACGATCACAATCAATCAA 319
      |||
Db 190 GATCGTGTATTATTCACGATCACAATCAATCAA 225

RESULT 14
  BJO74200      612 bp      mRNA      linear      EST 11-DEC-2001
  LOCUS      BJO74200 NIBB Mochli normalized Xenopus tailbud library Xenopus
  DEFINITION      laevis cDNA clone XL092h02 5', mRNA sequence.
  ACCESSION      BJO74200
  VERSION      BJO74200.1 GI:17504389
  KEYWORDS      EST.
  SOURCE      African clawed frog.
  ORGANISM      Xenopus laevis
  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
  Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
  Xenopodinae; Xenopus.
  REFERENCE      1 (bases 1 to 612)
  AUTHORS      Kitayama,A., Terasaka,C., Mochli,M., Ueno,N., Shin-I,T. and Kohara
  Y.
  TITLE      Expressed genes in X. laevis embryo
  JOURNAL      Unpublished (2001)
  COMMENT      Contact: Tadasu Shin-I
  National Institute of Genetics
  1111 Yata, Mishima, Shizuoka 411-8540, Japan
  Tel: 81-559-81-6856
  Fax: 81-559-81-6855
  Email: tshin@genes.nig.ac.jp.
  Location/Qualifiers
  1..612
    /organism="Xenopus laevis"
    /db_xref="taxon:8355"
    /clone="XL092h02"
    /clone_1id="NIBB Mochli normalized Xenopus tailbud
    library"
    /tissue_type="whole embryo"
    /dev_stage="stage 25"
BASE COUNT      191 a 144 c 133 g 144 t
ORIGIN
Query Match      0.3%; Score 33; DB 10; Length 612;
Best Local Similarity 100.0%; Pred. No. 0.00088;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8330 CATTATAGAGAGAAATATCATCATTTATTCACGG 8362
      |||
Db 104 CATTATAGAGAGAAATATCATCATTTATTCACGG 136

RESULT 15
  BJO86865/c      533 bp      mRNA      linear      EST 12-DEC-2001
  LOCUS      BJO86865

```

```

DEFINITION      BJO86865 NIBB Mochli normalized Xenopus tailbud library Xenopus
  laevis cDNA clone XL054i24 3', mRNA sequence.
ACCESSION      BJO86865
  BJO86865.1 GI:17583439
VERSION      EST.
KEYWORDS      African clawed frog.
SOURCE      Xenopus laevis
  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
  Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
  Xenopodinae; Xenopus.
  REFERENCE      1 (bases 1 to 533)
  AUTHORS      Kitayama,A., Terasaka,C., Mochli,M., Ueno,N., Shin-I,T. and Kohara
  Y.
  TITLE      Expressed genes in X. laevis embryo
  JOURNAL      Unpublished (2001)
  COMMENT      Contact: Tadasu Shin-I
  National Institute of Genetics
  1111 Yata, Mishima, Shizuoka 411-8540, Japan
  Tel: 81-559-81-6856
  Fax: 81-559-81-6855
  Email: tshin@genes.nig.ac.jp.
  Location/Qualifiers
  1..533
    /organism="Xenopus laevis"
    /db_xref="taxon:8355"
    /clone="XL054i24"
    /clone_1id="NIBB Mochli normalized Xenopus tailbud
    library"
    /tissue_type="whole embryo"
    /dev_stage="stage 25"
BASE COUNT      185 a 97 c 85 g 154 t 12 others
ORIGIN
Query Match      0.3%; Score 28; DB 10; Length 533;
Best Local Similarity 100.0%; Pred. No. 0.2;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9187 GTCACGCTCATTTTATATGTTATGTC 9214
      |||
Db 414 GTCACGCTCATTTTATATGTTATGTC 387

Search completed: October 7, 2002, 13:02:49
Job time : 6584 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 7, 2002, 09:33:18 ; Search time 888 Seconds
(without alignments)
18611.507 Million cell updates/sec

Title: US-09-150-867-2

Perfect score: 9626
Sequence: 1 gattccgcgagtcgcatag.....ttaaaaaaaaaacgcgaattc 9626

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

N_Geneseq_032802:*

1: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9626	100.0	9626	20	AA126819
2	23	0.2	7684	20	AA12953
3	22	0.2	15101	23	ABL08140
4	22	0.2	22191	20	AA120600
5	21	0.2	21	20	AA126825
6	21	0.2	692	23	ABL03005
7	21	0.2	2553	23	ABL26462
8	21	0.2	2751	23	ABL03004
9	21	0.2	3031	23	AA1273906

10	21	0.2	3054	23	AA1274043	DNA encoding novel
C 11	21	0.2	4117	23	ABL27156	Drosophila melanog
C 12	21	0.2	4699	23	ABL12872	Drosophila melanog
C 13	21	0.2	7545	23	AA123907	DNA encoding novel
14	21	0.2	7971	21	AA124473	Human immune/haema
15	21	0.2	8617	22	AA125052	Human polynucleoti
16	21	0.2	9018	24	AA121167	Human gene regulat
17	21	0.2	10564	16	AA103854	Human fibrinogen g
18	21	0.2	17141	24	AA126486	Bovine gamma globl
19	21	0.2	18997	24	ABL132571	Human immune syste
20	20	0.2	754	20	AA126086	Human gene express
21	20	0.2	951	21	AA125054	Arabidopsis thalia
22	22	0.2	1143	21	AA125057	Arabidopsis thalia
23	20	0.2	1198	20	AA129287	Human regulatory p
24	20	0.2	1202	20	AA128657	Human prostate tum
25	20	0.2	1205	21	AA123267	Human colorectal ca
26	20	0.2	1215	21	AA129069	Human colon cancer
27	20	0.2	1440	23	AA127068	DNA encoding novel
28	20	0.2	2097	14	AA125216	Coriolus hirsutus
C 29	20	0.2	2336	22	AA129791	S cerevisiae apoc
30	20	0.2	2391	23	ABL08025	Drosophila melanog
31	20	0.2	2580	22	AA129182	Human secreted pro
32	20	0.2	3403	23	ABL26975	Drosophila melanog
33	20	0.2	3640	24	AA129915	Polynucleotide enc
34	20	0.2	4061	22	AA158478	Human polynucleoti
35	20	0.2	4150	22	AA129122	Human polynucleoti
C 36	20	0.2	4176	22	AA129905	Human polynucleoti
C 37	20	0.2	4176	22	AA129905	Human polynucleoti
C 38	20	0.2	4176	22	AA160264	Human polynucleoti
39	20	0.2	4231	23	AA129707	DNA encoding novel
40	20	0.2	4258	22	AA129121	Human polynucleoti
41	20	0.2	4391	23	ABL08024	Drosophila melanog
42	20	0.2	5875	24	ABL22288	Human immune syste
C 43	20	0.2	5981	24	ABL33762	Human immune syste
44	20	0.2	6012	24	AA121275	Human gene regulat
45	20	0.2	6222	23	ABL26974	Drosophila melanog

ALIGNMENTS

RESULT 1	AA126819	
ID	AA126819	standard; DNA: 9626 BP.
XX	AA126819:	
AC	XX	
XX	XX	
DT	22-JUN-1999	(first entry)
DE	Nucleotide sequence of centromere-associated protein-E (CENP-E).	
XX	CENP-E; centromere-associated protein-E; ATPase activity;	
KW	plus end-directed microtubule motor activity; chromosome congression;	
KW	microtubule binding activity; chromosome movement; mitosis;	
KW	cell proliferation; tumor; metastasis; vascular malfunction;	
KW	inflammatory disease; immune disease; angiogenesis; hypertension;	
KW	stenosis; fungal infection; selective herbicide; fungicide;	
KW	insecticide; plant growth regulator; activator; cancer cell marker; ss.	
XX		
OS	Xenopus sp.	
XX		
PN	W09913061-A1.	
PD	18-MAR-1999.	
XX		
PF	10-SEP-1998;	98MO-US19231.
XX		
PR	11-SEP-1997;	97US-0058645.
XX		
PA	(REGC) UNIV CALIFORNIA.	
XX		
PI	Cleveland DW, Goldstein LSB, Sakowicz R, Wood KW;	
XX		

DR WPI: 1999-229233/19.
DR P-PSDB: AAY01632.
XX Centromere-associated protein-E and related nucleic acid
XX
XX
PS Claim 13; Page 67-73; 77pp: English.

CC The present sequence encodes CENP-E (centromere-associated protein-E)
CC of Xenopus. The protein has at least one of plus end-directed microtubule
CC motor activity, ATPase (adenosine triphosphatase) activity and
CC microtubule binding activity. CENP-E is the motor that powers chromosome
CC movement toward microtubule plus ends and is essential for congression
CC of chromosomes during mitosis. Modulators of CENP-E can thus control
CC cell proliferation. Agents that modulate CENP-E activity are lead
CC therapeutic, biologic, agricultural and diagnostic agents, e.g. for treatment
CC of unwanted cell proliferation (typical of many examples are tumors and
CC metastases; vascular malfunction; inflammatory and immune diseases;
CC angiogenesis; hypertension; restenosis; and fungal infections), also as
CC plant-protection agents (selective herbicides, fungicides and
CC insecticides) and plant growth regulators or activators for improving
CC yields. CENP-E is also a diagnostic marker for dividing cells, including
CC cancer cells.
XX

Sequence 9626 BP: 3540 A: 1692 C: 2104 G: 2290 T: 0 other:

Query Match 100.0%; Score 9626; DB 20; Length 9626;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 9626; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCGGAGTGGATAGGCTACTCGGCGAGGAAATTCAACTGCTTATAGAGAAA 60
DB 1 GAATTCGGAGTGGATAGGCTACTCGGCGAGGAAATTCAACTGCTTATAGAGAAA 60
QY 61 CTTGACCGCCGCCCAAAAGGACTAAAGTACAGACAGAGGAGCGGTGTCCGTACCGAT 120
DB 61 CTTGACCGCCGCCCAAAAGGACTAAAGTACAGACAGAGGAGCGGTGTCCGTACCGAT 120
QY 121 TTCTCCACTATGCTGCTCAAAATGTCGCGAGAGATGCATTAAAGTGTGTGAGGGT 180
DB 121 TTCTCCACTATGCTGCTCAAAATGTCGCGAGAGATGCATTAAAGTGTGTGAGGGT 180
QY 181 TCGGCGCTTATACAGAGAACAGGAGATCAAGCCACTTCATATGGAAGGCTGGAAA 240
DB 181 TCGGCGCTTATACAGAGAACAGGAGATCAAGCCACTTCATATGGAAGGCTGGAAA 240
QY 241 CAACACCATTTCCCAAGTGTGATGCGACAAAGTCTTCAATTCGATCGTATTTAATTC 300
DB 241 CAACACCATTTCCCAAGTGTGATGCGACAAAGTCTTCAATTCGATCGTATTTAATTC 300
QY 301 TCACGATCAACAAGTCAAAATTTACCAAGAAATAGCAGTACCTATCATACGATCAGCTTT 360
DB 301 TCACGATCAACAAGTCAAAATTTACCAAGAAATAGCAGTACCTATCATACGATCAGCTTT 360
QY 361 GCAGGAGTATATGCGACATAATTTGATAGCAGACAGACATCTTCAGGACAAACGTACAC 420
DB 361 GCAGGAGTATATGCGACATAATTTGATAGCAGACAGACATCTTCAGGACAAACGTACAC 420
QY 421 AATGATGGGAACACCAATTCATTTGGGCTATATACCCCAAGCCATACAGAAATTTTAA 480
DB 421 AATGATGGGAACACCAATTCATTTGGGCTATATACCCCAAGCCATACAGAAATTTTAA 480
QY 481 AATATTCAGAGATACCGAACAGAGAGTTCTTCTAAGAGTTTCTTATATGAGAAATTTA 540
DB 481 AATATTCAGAGATACCGAACAGAGAGTTCTTCTAAGAGTTTCTTATATGAGAAATTTA 540
QY 541 CAATCAAACTGTGAAAGACCTACTGTGTGATGACAGAGAAAGAAAGCCCTTGAAATTCG 600
DB 541 CAATCAAACTGTGAAAGACCTACTGTGTGATGACAGAGAAAGAAAGCCCTTGAAATTCG 600
QY 601 CGAGGATTTTATAGAAACGTGTATGCTGACCTGCTGAAACACTTGTAAATGTTCC 660
DB 601 CGAGGATTTTATAGAAACGTGTATGCTGACCTGCTGAAACACTTGTAAATGTTCC 660

QY 661 TGAACATGTAATACAGTGTGATCAAAAAAGGTGAAAAAACAGACATTATGACAGACTAA 720
DB 661 TGAACATGTAATACAGTGTGATCAAAAAAGGTGAAAAAACAGACATTATGACAGACTAA 720
QY 721 AATGAATGATCATATGATGCTGCTCACATACATATTTAGAAATATTTGTAAGCCGACA 780
DB 721 AATGAATGATCATATGATGCTGCTCACATACATATTTAGAAATATTTGTAAGCCGACA 780
QY 781 CAGAAATGATCCACAAATTCAGAGAACTGTGATGAGAGCTGATGATGATTCACACTTGA 840
DB 781 CAGAAATGATCCACAAATTCAGAGAACTGTGATGAGAGCTGATGATGATTCACACTTGA 840
QY 841 TTGTGTAGATCTGCTGCGACAGTGAAGAGACAGCCAACTGAGCTGAGGTGTGAGACT 900
DB 841 TTGTGTAGATCTGCTGCGACAGTGAAGAGACAGCCAACTGAGCTGAGGTGTGAGACT 900
QY 901 TAAGGAAGCTGTGACATCAACCGCAGCTGTGTTATCTTGGACGAGTATTAACAGCT 960
DB 901 TAAGGAAGCTGTGACATCAACCGCAGCTGTGTTATCTTGGACGAGTATTAACAGCT 960
QY 961 TAGCAGCGCCAGGCTGTGATTTATAACTACAGAGACAGCAACCTCACAGAAATTC 1020
DB 961 TAGCAGCGCCAGGCTGTGATTTATAACTACAGAGACAGCAACCTCACAGAAATTC 1020
QY 1021 CCAAAATTCATTTGGGAGGAAATGCTTAAACGCTTATATTTGACAAATTTAGCCAGTTTC 1080
DB 1021 CCAAAATTCATTTGGGAGGAAATGCTTAAACGCTTATATTTGACAAATTTAGCCAGTTTC 1080
QY 1081 TTTTGTAGAGCTTAAGTACACCTTCAGTTTSCAGTACTGCAAAACATGTGAGAAATAC 1140
DB 1081 TTTTGTAGAGCTTAAGTACACCTTCAGTTTSCAGTACTGCAAAACATGTGAGAAATAC 1140
QY 1141 TCCCATGTTAATGAGTCTCTGATGATGAACGCTGTCTAAAAAGTACAGAAAGAAAT 1200
DB 1141 TCCCATGTTAATGAGTCTCTGATGATGAACGCTGTCTAAAAAGTACAGAAAGAAAT 1200
QY 1201 CTTGATTTAAAGAAACATTTAGAGATTTAGAGTATGATGATGATGATGATGATGATGAT 1260
DB 1201 CTTGATTTAAAGAAACATTTAGAGATTTAGAGTATGATGATGATGATGATGATGATGAT 1260
QY 1261 AATGCTTAAAGAGACATACACAGTTGCTAGCTGAATCAACACATACACAAAGAGAG 1320
DB 1261 AATGCTTAAAGAGACATACACAGTTGCTAGCTGAATCAACACATACACAAAGAGAG 1320
QY 1321 AGAATGATGAATATGAGCTTGCACAAATATGTTGCTTCAATCCCAAGATTCACACA 1380
DB 1321 AGAATGATGAATATGAGCTTGCACAAATATGTTGCTTCAATCCCAAGATTCACACA 1380
QY 1381 GGACCAAGGGTCAAAAGCAAAAGAGATGAGTACGTGGGCGCAGAGAAATCCAAATAG 1440
DB 1381 GGACCAAGGGTCAAAAGCAAAAGAGATGAGTACGTGGGCGCAGAGAAATCCAAATAG 1440
QY 1441 TTTACATGCTTCTGCTGTTCTGACTTTGATATGCTATCCAGATTTACCTGGCAATTTAG 1500
DB 1441 TTTACATGCTTCTGCTGTTCTGACTTTGATATGCTATCCAGATTTACCTGGCAATTTAG 1500
QY 1501 CAAGAAGCAAAAGTCTCTGACATGCTTCATTTCCGAATGATGAGTCTGTTGTGAT 1560
DB 1501 CAAGAAGCAAAAGTCTCTGACATGCTTCATTTCCGAATGATGAGTCTGTTGTGAT 1560
QY 1561 AGAGTTTGTGATTTGATGATGAGGCTCTCCATGATGAGACAGCAATGATGATGATGAT 1620
DB 1561 AGAGTTTGTGATTTGATGATGAGGCTCTCCATGATGAGACAGCAATGATGATGATGAT 1620
QY 1621 ATGGAATTTAGCCAGTAAATTAATCAAGAGAAAGACATCACTTCAATCAATGAT 1680
DB 1621 ATGGAATTTAGCCAGTAAATTAATCAAGAGAAAGACATCACTTCAATCAATGAT 1680
QY 1681 AGACTTTGACAGATTTCTGACAGTGTTCAGTTTCAATGATTTCTTAAAGAAACACAGCT 1740
DB 1681 AGACTTTGACAGATTTCTGACAGTGTTCAGTTTCAATGATTTCTTAAAGAAACACAGCT 1740
QY 1741 ACAATACCTCCCAAGACCTGTGATATGGCTGATGACAGAAAGCTTTTGTGAAA 1800

```
Db 1741 ACACATACCTCCCAAGACCTGCTGATATGCTGATGACAGAAAAGCTCTTTGAAA 1800
Oy 1801 AGAGATCACAGCCTCCAGCAACAATACATGTCAAAAGAGAAAGAAAAGAACTGCT 1860
Db 1801 AGAGATCACAGCCTCCAGCAACAATACATGTCAAAAGAGAAAGAAAAGAACTGCT 1860
Oy 1861 ACAAAAGCTTCGAGCTCAAGATGACAACTGAGAGAGCACTTAGTGCTAAAGCTAA 1920
Db 1861 ACAAAAGCTTCGAGCTCAAGATGACAACTGAGAGAGCACTTAGTGCTAAAGCTAA 1920
Oy 1921 TCTAGAGATGGTTTACAACTCGAGAGAGCATTCATTAATGCTGAAGTCCAAACAGATG 1980
Db 1921 TCTAGAGATGGTTTACAACTCGAGAGAGCATTCATTAATGCTGAAGTCCAAACAGATG 1980
Oy 1981 TGAAGAAGAGTGTGAGAGAAAGAAATGCTCACTCTTGAGAGACTGCTGGTTACAAATGCAATC 2040
Db 1981 TGAAGAAGAGTGTGAGAGAAAGAAATGCTCACTCTTGAGAGACTGCTGGTTACAAATGCAATC 2040
Oy 2041 AAACAGTACCTACAGATAGTTCTGTTGATGCTAAAGCTCAGAGAGCTCCCATGATGA 2100
Db 2041 AAACAGTACCTACAGATAGTTCTGTTGATGCTAAAGCTCAGAGAGCTCCCATGATGA 2100
Oy 2101 GTGATAGAACACAGAAAAAATGCTGGAACAAAGATCGTTGATTTAGAAAGATTTATGA 2160
Db 2101 GTGATAGAACACAGAAAAAATGCTGGAACAAAGATCGTTGATTTAGAAAGATTTATGA 2160
Oy 2161 AAACCTTAAACAAGAAAGTGAAGATGATTAACAAAAATCTTCTGAGCAAGATTTATGA 2220
Db 2161 AAACCTTAAACAAGAAAGTGAAGATGATTAACAAAAATCTTCTGAGCAAGATTTATGA 2220
Oy 2221 GAGTATAGCTATGTGAAGCTATATATGCGAAGAAAGCAATGCACTGAGAGAACTGGC 2280
Db 2221 GAGTATAGCTATGTGAAGCTATATATGCGAAGAAAGCAATGCACTGAGAGAACTGGC 2280
Oy 2281 ACTTATGCGAGATTAATTTTGACAAATATTAATCTAGAGATGAAGCTCTAAAAAGGAAAT 2340
Db 2281 ACTTATGCGAGATTAATTTTGACAAATATTAATCTAGAGATGAAGCTCTAAAAAGGAAAT 2340
Oy 2341 TGCAGATCTGGAGACCTTCACTCAAGAGAAATCAAAAACCAATGAGTTGAAATCTGGA 2400
Db 2341 TGCAGATCTGGAGACCTTCACTCAAGAGAAATCAAAAACCAATGAGTTGAAATCTGGA 2400
Oy 2401 GAAGAAACTCAAAAAGAACACAGAGCAACTAATCCATGATGGAGTTTAAAGA 2460
Db 2401 GAAGAAACTCAAAAAGAACACAGAGCAACTAATCCATGATGGAGTTTAAAGA 2460
Oy 2461 ATTAGTTGAAAAATGAGAGATGTACAAATCTTGAAGAGATCTAGAAACTAAAC 2520
Db 2461 ATTAGTTGAAAAATGAGAGATGTACAAATCTTGAAGAGATCTAGAAACTAAAC 2520
Oy 2521 AAAACTTCTGAAGAGACGAGAAATCAACTTGCAGAAATTAAGAAACGGCGAGATTA 2580
Db 2521 AAAACTTCTGAAGAGACGAGAAATCAACTTGCAGAAATTAAGAAACGGCGAGATTA 2580
Oy 2581 GCAGAAAAAGATGAGAAATTTTATGATCTCGGTTCCAGTGGGTGATGAGAAACTGCG 2640
Db 2581 GCAGAAAAAGATGAGAAATTTTATGATCTCGGTTCCAGTGGGTGATGAGAAACTGCG 2640
Oy 2641 TGAAGAAATCTTTCAACTGAAGCAATCTTCTGATGCTGAAGCTGACTCGGATGTC 2700
Db 2641 TGAAGAAATCTTTCAACTGAAGCAATCTTCTGATGCTGAAGCTGACTCGGATGTC 2700
Oy 2701 TCAGAGAGAAATGCTCTTCAGAGAGTGAAGATCTAGAGCTGAAGAGAAATGAGAGA 2760
Db 2701 TCAGAGAGAAATGCTCTTCAGAGAGTGAAGATCTAGAGCTGAAGAGAAATGAGAGA 2760
Oy 2761 CACATCAAACTGTGATCATCAAAAAGAAAGCTGCTTTGTTGAGAAAGCACTGGA 2820
Db 2761 CACATCAAACTGTGATCATCAAAAAGAAAGCTGCTTTGTTGAGAAAGCACTGGA 2820
Oy 2821 AACGAAAAATCAAACTCAAGAAAAATGAGAGCTGATTTGAGAAAGGTTGCAAAAGTGC 2880
Db 2821 AACGAAAAATCAAACTCAAGAAAAATGAGAGCTGATTTGAGAAAGGTTGCAAAAGTGC 2880
Db 2821 AACGAAAAATCAAACTCAAGAAAAATGAGAGCTGATTTGAGAAAGGTTGCAAAAGTGC 2880
Oy 2881 TTTTAAATGAGATTAACCTACTTAATGAGCTTCTGCGAGAGAAAGTCCCGAGATTTGCT 2940
Db 2881 TTTTAAATGAGATTAACCTACTTAATGAGCTTCTGCGAGAGAAAGTCCCGAGATTTGCT 2940
Oy 2941 TTCTGCTGTTGAATTTAGAGAAAAAGTTCGAGTTCTGAAGACAGCTTGACAAAGCAT 3000
Db 2941 TTCTGCTGTTGAATTTAGAGAAAAAGTTCGAGTTCTGAAGACAGCTTGACAAAGCAT 3000
Oy 3001 GGAAGAAAAAATGCTTGGAGAAATGAGTGAATGCACTTATCAGAAATCAAAATTTTGGC 3060
Db 3001 GGAAGAAAAAATGCTTGGAGAAATGAGTGAATGCACTTATCAGAAATCAAAATTTTGGC 3060
Oy 3061 AAATGAGTTGAATGCTTTAAAAATCAGATCAGCAAGGCTTCTGAAGAGATTAATGAT 3120
Db 3061 AAATGAGTTGAATGCTTTAAAAATCAGATCAGCAAGGCTTCTGAAGAGATTAATGAT 3120
Oy 3121 AAAGCAAGAGAGAAACATCTGCAATCTATTAAGCAAAACAGAGATTAATCAGAGGA 3180
Db 3121 AAAGCAAGAGAGAAACATCTGCAATCTATTAAGCAAAACAGAGATTAATCAGAGGA 3180
Oy 3181 GCAATCTGAGCAGATTTTACAACCTGACGAGAGTGAACACACACAGCTCAAAAGTGA 3240
Db 3181 GCAATCTGAGCAGATTTTACAACCTGACGAGAGTGAACACACACAGCTCAAAAGTGA 3240
Oy 3241 GCAGACTGAAGAGCAATCTTGAAGTGAAGAAATGCAATGATCTTTTGAAGAGTA 3300
Db 3241 GCAGACTGAAGAGCAATCTTGAAGTGAAGAAATGCAATGATCTTTTGAAGAGTA 3300
Oy 3301 TATCAGAAACAAAGTGAAGTGAAGACCTTTAAGAGAAATGGAACCTTAAAGGCAC 3360
Db 3301 TATCAGAAACAAAGTGAAGTGAAGACCTTTAAGAGAAATGGAACCTTAAAGGCAC 3360
Oy 3361 TATGAGTCTGTGGAAGTGAAGATTTGAGACAAACATGAGACTTGAAGAACTATAG 3420
Db 3361 TATGAGTCTGTGGAAGTGAAGATTTGAGACAAACATGAGACTTGAAGAACTATAG 3420
Oy 3421 GGATTAAGAGCAACCTGCTCATGAGAAAAAATCACTTTTCAAGCAATCCACATATAT 3480
Db 3421 GGATTAAGAGCAACCTGCTCATGAGAAAAAATCACTTTTCAAGCAATCCACATATAT 3480
Oy 3481 TCCGATTAACCTCTTTGAGACTCGCTTCCCTCAAAATTAATGTTGAAGGAACTGCA 3540
Db 3481 TCCGATTAACCTCTTTGAGACTCGCTTCCCTCAAAATTAATGTTGAAGGAACTGCA 3540
Oy 3541 AGACCCCATGAGAAATCAATGACTACCAAAATTAATAGCCCTTGCTACAGAAAGAACAA 3600
Db 3541 AGACCCCATGAGAAATCAATGACTACCAAAATTAATAGCCCTTGCTACAGAAAGAACAA 3600
Oy 3601 CATTAATGCTGTCTTAGAGACTGAGAAAGAAACAGTCTCAAGGAGCAAGTTATGATTTGA 3660
Db 3601 CATTAATGCTGTCTTAGAGACTGAGAAAGAAACAGTCTCAAGGAGCAAGTTATGATTTGA 3660
Oy 3661 CACTCAACTTCAAAAGTCTTCAAGCAACAAAGTATGAAAAAGTGTGATCTCCAGAACCCAA 3720
Db 3661 CACTCAACTTCAAAAGTCTTCAAGCAACAAAGTATGAAAAAGTGTGATCTCCAGAACCCAA 3720
Oy 3721 GCAAGACTTGAAGAGAGAGAGTTAACTGCTTTTGAAGATGGAATCACTGAAGGACA 3780
Db 3721 GCAAGACTTGAAGAGAGAGAGTTAACTGCTTTTGAAGATGGAATCACTGAAGGACA 3780
Oy 3781 CCTAATGACTCAGAGCTGTCTATAGAGAAATGCAAGCTGAGAAATCTGGAAGTTACAGA 3840
Db 3781 CCTAATGACTCAGAGCTGTCTATAGAGAAATGCAAGCTGAGAAATCTGGAAGTTACAGA 3840
Oy 3841 AAAACTCAAACTCTTCAAGAAAGATGAAAAATTTACTATAGAAAGGAATGACCTTCA 3900
Db 3841 AAAACTCAAACTCTTCAAGAAAGATGAAAAATTTACTATAGAAAGGAATGACCTTCA 3900
Oy 3901 GACCAACTTGAAGAGCTGAAAGAGAGATGATAGCTTAAAGCAAGACTTATAGGAAA 3960
Db 3901 GACCAACTTGAAGAGCTGAAAGAGAGATGATAGCTTAAAGCAAGACTTATAGGAAA 3960
```

```
Oy 3961 CATTGAGCAGTCATTTGCAACAAGATGAATTAAAGGCTGCCAGAGAGAGCTTAAGAGA 4020
Db 3961 CATTGAGCAGTCATTTGCAACAAGATGAATTAAAGGCTGCCAGAGAGAGCTTAAGAGA 4020
Oy 4021 ACAGAGCAATGGTGTGATGACTTTAGACAACACCTTTTGTGATGTCTGTAGAAATTC 4080
Db 4021 ACAGAGCAATGGTGTGATGACTTTAGACAACACCTTTTGTGATGTCTGTAGAAATTC 4080
Oy 4081 ATCAACCAACCATATGATGAGTTGCTAAACAGAGAAAAGGTGTCAATGGGTGAATTAATTC 4140
Db 4081 ATCAACCAACCATATGAGTTGCTAAACAGAGAAAAGGTGTCAATGGGTGAATTAATTC 4140
Oy 4141 GTTACAAAGTGAATTCGTGGCTGGTGAAGATGAGCTTCAACATCTTGTAAAGCATTT 4200
Db 4141 GTTACAAAGTGAATTCGTGGCTGGTGAAGATGAGCTTCAACATCTTGTAAAGCATTT 4200
Oy 4201 AGTTTCAGAACTGAGACTACTTGTGTCTCATGTAAATCTGTGAGAGAGAAAACCTTGA 4260
Db 4201 AGTTTCAGAACTGAGACTACTTGTGTCTCATGTAAATCTGTGAGAGAGAAAACCTTGA 4260
Oy 4261 AATCACAAAAAACTCAATGGCTTGAAGAGAGATTTGGCAATTTGGAAGAACGA 4320
Db 4261 AATCACAAAAAACTCAATGGCTTGAAGAGAGATTTGGCAATTTGGAAGAACGA 4320
Oy 4321 AGTTTGAAGCTCATGTGAGAAATCTAAAGAGACACAAATTAAGCTCAAGAACAGC 4380
Db 4321 AGTTTGAAGCTCATGTGAGAAATCTAAAGAGACACAAATTAAGCTCAAGAACAGC 4380
Oy 4381 AGAGGATATTTCTAGTAAAGAAAATCAATTCAGTTTGAAGAGGTTCAGTGTTCACA 4440
Db 4381 AGAGGATATTTCTAGTAAAGAAAATCAATTCAGTTTGAAGAGGTTCAGTGTTCACA 4440
Oy 4441 GAAAGCTGATAGATGAGATGAGGTCTGTAAGACAGCTAAAGGACAGAGAAAGAGCT 4500
Db 4441 GAAAGCTGATAGATGAGATGAGGTCTGTAAGACAGCTAAAGGACAGAGAAAGAGCT 4500
Oy 4501 GGAATTAAGAGATAGATATTTTGAATCTTACAACTGCAACACCAATTTAGTTGA 4560
Db 4501 GGAATTAAGAGATAGATATTTTGAATCTTACAACTGCAACACCAATTTAGTTGA 4560
Oy 4561 GGGAAAAATTTGAAATCTCATTTGCAAGCTGACATGAGAGAAAGACAGCATTTGCGCGTTC 4620
Db 4561 GGGAAAAATTTGAAATCTCATTTGCAAGCTGACATGAGAGAAAGACAGCATTTGCGCGTTC 4620
Oy 4621 TGAAGAAATGGAGATTAAGATTTCTTGAGAAACCTTGAGCAATCAGATTTACTGGA 4680
Db 4621 TGAAGAAATGGAGATTAAGATTTCTTGAGAAACCTTGAGCAATCAGATTTACTGGA 4680
Oy 4681 AAGATTTGCAAGAGAAAGCTGGAACCTGTCTAACAACTTGAATCTCTGAGAAAGAT 4740
Db 4681 AAGATTTGCAAGAGAAAGCTGGAACCTGTCTAACAACTTGAATCTCTGAGAAAGAT 4740
Oy 4741 GGAGAGCTGCTTCTATTTAAAGATGACCTGCAACAGAACTGGAAGTTCGTGAGTGA 4800
Db 4741 GGAGAGCTGCTTCTATTTAAAGATGACCTGCAACAGAACTGGAAGTTCGTGAGTGA 4800
Oy 4801 AAACATTTCTTAAAGAGATTTGACACAACTTGAAGCATCATCTCAATCTCAAGC 4860
Db 4801 AAACATTTCTTAAAGAGATTTGACACAACTTGAAGCATCATCTCAATCTCAAGC 4860
Oy 4861 TGCAGCTGCAAGAAACACAGCAGAGCTACAGTTGGCTTAAGAAATCTTGCAATCCCTCTTC 4920
Db 4861 TGCAGCTGCAAGAAACACAGCAGAGCTACAGTTGGCTTAAGAAATCTTGCAATCCCTCTTC 4920
Oy 4921 TGACAAATGTCATACTACAGAGAAAGAACTCTGCAAGATTTGTGATCTCTTGGGA 4980
Db 4921 TGACAAATGTCATACTACAGAGAAAGAACTCTGCAAGATTTGTGATCTCTTGGGA 4980
Oy 4981 AGAAAGATATTTGTTATTAATCTGAAGATTTGATCAAAAAAATTAAGAGAGAGAAATTT 5040
Db 4981 AGAAAGATATTTGTTATTAATCTGAAGATTTGATCAAAAAAATTAAGAGAGAGAAATTT 5040
```

```
Oy 5041 ACTACATGAAGAAATGAACTTGAACCAAGCTCAGCTTAAAGTGCAGGTGGAACA 5100
Db 5041 ACTACATGAAGAAATGAACTTGAACCAAGCTCAGCTTAAAGTGCAGGTGGAACA 5100
Oy 5101 TCTGATGAAGATATGATGATGCAATGCAAGTCTCACTTGTGCTGGAGCATGGAACA 5160
Db 5101 TCTGATGAAGATATGATGATGCAATGCAAGTCTCACTTGTGCTGGAGCATGGAACA 5160
Oy 5161 TGATCTGAACACACACTTCTTGTCTTAAACAGCAGATGCAATGTTACTCAAGAAA 5220
Db 5161 TGATCTGAACACACACTTCTTGTCTTAAACAGCAGATGCAATGTTACTCAAGAAA 5220
Oy 5221 GAAAGAGCTGCAACAAACCCATGAACTTAAAGAGCTGAGGTGACCATTTAAAGAGAA 5280
Db 5221 GAAAGAGCTGCAACAAACCCATGAACTTAAAGAGCTGAGGTGACCATTTAAAGAGAA 5280
Oy 5281 TATTGAATTTGGTGTGAATTTTAAATTAAGAGCAGCAAAAGCAGTAAAGAGCAATG 5340
Db 5281 TATTGAATTTGGTGTGAATTTTAAATTAAGAGCAGCAAAAGCAGTAAAGAGCAATG 5340
Oy 5341 TCTGCTAAATGAGATTAAGAACTTGAACCAAGCAGCAGACAGCTTCAATGTGATAGA 5400
Db 5341 TCTGCTAAATGAGATTAAGAACTTGAACCAAGCAGCAGACAGCTTCAATGTGATAGA 5400
Oy 5401 AGACCTTAATGAAGAACTTAAAGATTAAGAGTCAAGGCTGGAACCTTAAAGAAATCTGA 5460
Db 5401 AGACCTTAATGAAGAACTTAAAGATTAAGAGTCAAGGCTGGAACCTTAAAGAAATCTGA 5460
Oy 5461 ACAAAAGTAAATTAATGTAACCAAGAAATGGAATGCTTATGTCTGGAATGAGAGAAAT 5520
Db 5461 ACAAAAGTAAATTAATGTAACCAAGAAATGGAATGCTTATGTCTGGAATGAGAGAAAT 5520
Oy 5521 GAAAAATGACCAAGAGAGCTGTAATGCTGAGAGAGCAGCTGCAAGAGAGCTTAAGGA 5580
Db 5521 GAAAAATGACCAAGAGAGCTGTAATGCTGAGAGAGCAGCTGCAAGAGAGCTTAAGGA 5580
Oy 5581 GAGTGTGAATGATGCTCATTGAAACCTCAAGATGATCTTAAGAAAGGCTCAAGAGATTCCA 5640
Db 5581 GAGTGTGAATGATGCTCATTGAAACCTCAAGATGATCTTAAGAAAGGCTCAAGAGATTCCA 5640
Oy 5641 GCAGCGAAAGATTAAGTTCAAGAACTGACCTCCAGATTTCTGCTGACAGAAAGAT 5700
Db 5641 GCAGCGAAAGATTAAGTTCAAGAACTGACCTCCAGATTTCTGCTGACAGAAAGAT 5700
Oy 5701 CTCTCTTGGAAAAATCAATGCTTTATTAATGTTGCAACTGTGAAAGAACTTAAGCGA 5760
Db 5701 CTCTCTTGGAAAAATCAATGCTTTATTAATGTTGCAACTGTGAAAGAACTTAAGCGA 5760
Oy 5761 AAGAGATGACCTGAACCAAGTCTAAGCAACACCTGTTCTCAGAAATTTGAACTCTTAAGCT 5820
Db 5761 AAGAGATGACCTGAACCAAGTCTAAGCAACACCTGTTCTCAGAAATTTGAACTCTTAAGCT 5820
Oy 5821 GTCTTTAAAGAAAGAAATTTGCAATTTGGAACAGCAGAGAGCAAAAGCTGATCTGC 5880
Db 5821 GTCTTTAAAGAAAGAAATTTGCAATTTGGAACAGCAGAGAGCAAAAGCTGATCTGC 5880
Oy 5881 CAGGAAAAACAATAGATATCAAGAGAAAGATTCAAATTTAAGAAACAGTACTTCAACA 5940
Db 5881 CAGGAAAAACAATAGATATCAAGAGAAAGATTCAAATTTAAGAAACAGTACTTCAACA 5940
Oy 5941 AGCCACCAATTTAAAGAACTCTTTATGAAGAGAGAGTCTTATVCCAGTGAAGAGCA 6000
Db 5941 AGCCACCAATTTAAAGAACTCTTTATGAAGAGAGAGTCTTATVCCAGTGAAGAGCA 6000
Oy 6001 ACTGCTTTGAACACAGAAACCTTGAAGAAACATTAAGAGCAAAAGCTTGGCATTTGGG 6060
Db 6001 ACTGCTTTGAACACAGAAACCTTGAAGAAACATTAAGAGCAAAAGCTTGGCATTTGGG 6060
Oy 6061 TAAATGAGAGAGAGAGATGAGTGCCTTAATAGTCTTTACAGAAAAAGAT 6120
Db 6061 TAAATGAGAGAGAGAGATGAGTGCCTTAATAGTCTTTACAGAAAAAGAT 6120
Oy 6121 GTCTTCTTGAAGAGAGATGCAATGAAATGTTACTTCTTAAAGAGAGGTGAGGTGA 6180
```

```
|||||
Db 6121 GTCTCTCTAGAGAACAGATCAATGAAATGTTACTACTCTAAAGAAAGGTGAGGTGA 6180
Qy 6181 AAAAGAACCTTCTACCTTCAGAGACCTTCAAGACAGACTCTTCCCAAGATGAGA 6240
Db 6181 AAAAGAACCTTCTACCTTCAGAGACCTTCAAGACAGACTCTTCCCAAGATGAGA 6240
Qy 6241 GCTCAGAGAGCTTTAAAGACCAAGATTTGCAGTTGGAAAGAGCCGAGAAAGAGATAG 6300
Db 6241 GCTCAGAGAGCTTTAAAGACCAAGATTTGCAGTTGGAAAGAGCCGAGAAAGAGATAG 6300
Qy 6301 TGAAGCTACTAATGAATTAAGAATCTCAGTGTAGATCTCTCTCTAGAGAGAGAT 6360
Db 6301 TGAAGCTACTAATGAATTAAGAATCTCAGTGTAGATCTCTCTCTAGAGAGAGAT 6360
Qy 6361 TCTTCAGATATGCTAGCACTTTTGAATGAAGCTGTAGCGAAAAGSAAACCTTCGCTTC 6420
Db 6361 TCTTCAGATATGCTAGCACTTTTGAATGAAGCTGTAGCGAAAAGSAAACCTTCGCTTC 6420
Qy 6421 GAAGCAGCACTGGTTTCAGAAATGGAGCACTATCAGTGCATTAAAGATAGAGACCA 6480
Db 6421 GAAGCAGCACTGGTTTCAGAAATGGAGCACTATCAGTGCATTAAAGATAGAGACCA 6480
Qy 6481 TGCATTTGCGCAATCTTAAAGCGCAAAAGATGAAGCTGTAAATTAATTAAGCAGTCTCGC 6540
Db 6481 TGCATTTGCGCAATCTTAAAGCGCAAAAGATGAAGCTGTAAATTAATTAAGCAGTCTCGC 6540
Qy 6541 TGAAGAAATTAAGATCTGACAAAGAGATGATGAATTCAGAGATTCAAAGAAATCTT 6600
Db 6541 TGAAGAAATTAAGATCTGACAAAGAGATGATGAATTCAGAGATTCAAAGAAATCTT 6600
Qy 6601 GCAGAAACACTCTTCCCATTAAGTGAAGATTAATGATTAAGCTCAACTCAAT 6660
Db 6601 GCAGAAACACTCTTCCCATTAAGTGAAGATTAATGATTAAGCTCAACTCAAT 6660
Qy 6661 GCTCAAGCAACAGAAAGAAAGACATCAACAAACTTGACAGAAATTAAGAAAGTGA 6720
Db 6661 GCTCAAGCAACAGAAAGAAAGACATCAACAAACTTGACAGAAATTAAGAAAGTGA 6720
Qy 6721 TGAAGTATTTGCAACACTTATCTCTTAAAGAAACAGCTGGACCAAAATACAGATGAGCT 6780
Db 6721 TGAAGTATTTGCAACACTTATCTCTTAAAGAAACAGCTGGACCAAAATACAGATGAGCT 6780
Qy 6781 AAGGAATGAAAAGCTCAGAAACTATGAACTCTCGGAAAAGATGATTCATGGAAAAAGA 6840
Db 6781 AAGGAATGAAAAGCTCAGAAACTATGAACTCTCGGAAAAGATGATTCATGGAAAAAGA 6840
Qy 6841 AATCTCAGTCTGCTGTTTATGAGAAAGAGCCCTCAGCAGAAAGATGATGTTGCGAGA 6900
Db 6841 AATCTCAGTCTGCTGTTTATGAGAAAGAGCCCTCAGCAGAAAGATGATGTTGCGAGA 6900
Qy 6901 ACSTATGATATTAATCTGAGAGCAAAACCAAGAAATACAGAGCTGATGGAAAAATCTC 6960
Db 6901 ACSTATGATATTAATCTGAGAGCAAAACCAAGAAATACAGAGCTGATGGAAAAATCTC 6960
Qy 6961 CGCTGTGTATTCAGAGCAACACTTTGCTCAGCAGTCTCTAGTGAAGCTTCAAAAGGA 7020
Db 6961 CGCTGTGTATTCAGAGCAACACTTTGCTCAGCAGTCTCTAGTGAAGCTTCAAAAGGA 7020
Qy 7021 AACTGAAAGCAACAAACATTCATGATTAATTAAGAGATCTCTGATCCAGCCCTC 7080
Db 7021 AACTGAAAGCAACAAACATTCATGATTAATTAAGAGATCTCTGATCCAGCCCTC 7080
Qy 7081 CAGATCCTTTGGAGCTTGCAGAACTGAGCATGTTAAGCTAAATCTCAACTGCAGACCT 7140
Db 7081 CAGATCCTTTGGAGCTTGCAGAACTGAGCATGTTAAGCTAAATCTCAACTGCAGACCT 7140
Qy 7141 TCTGAAACAATTTAAGTTTATACGAACTGCTGCAGTCAAAAGAAATATAGCTTGAT 7200
Db 7141 TCTGAAACAATTTAAGTTTATACGAACTGCTGCAGTCAAAAGAAATATAGCTTGAT 7200
Qy 7201 CAAGATTTATGAGAAAGACCTTCTGCTGAGCAAAAGAGCATGATGAGCTGCAGTCA 7260
Db 7201 CAAGATTTATGAGAAAGACCTTCTGCTGAGCAAAAGAGCATGATGAGCTGCAGTCA 7260
|||||
Db 7201 CAAGATTTATGAGAAAGACCTTCTGCTGAGCAAAAGAGCATGATGAGCTGCAGTCA 7260
Qy 7261 ACTGCAGTGTTTGAGAGCAGCAGGCAAGAAATGCTCGGATTTCTGATCTAGAGACTCA 7320
Db 7261 ACTGCAGTGTTTGAGAGCAGCAGGCAAGAAATGCTCGGATTTCTGATCTAGAGACTCA 7320
Qy 7321 GTTCTGTGAATTAATTAATCTTGAATGATTAATCTTTTAAAAAAGCAATTAATTCAGAG 7380
Db 7321 GTTCTGTGAATTAATTAATCTTGAATGATTAATCTTTTAAAAAAGCAATTAATTCAGAG 7380
Qy 7381 TGTCCAGATGACTTTTTCAGAGGTGAGATTTCTTAATCAACTAGATCAACACTGCA 7440
Db 7381 TGTCCAGATGACTTTTTCAGAGGTGAGATTTCTTAATCAACTAGATCAACACTGCA 7440
Qy 7441 AGAAGACCTTGAGACAAAGAAAGCTTTATGCAATGAGTGTGGAGAAATTTGAGATCTGCA 7500
Db 7441 AGAAGACCTTGAGACAAAGAAAGCTTTATGCAATGAGTGTGGAGAAATTTGAGATCTGCA 7500
Qy 7501 CGTGCATGCTTAAGAACTCAGTGAAGCATGCAACAGGAAATAGCGCATGCTTCTAC 7560
Db 7501 CGTGCATGCTTAAGAACTCAGTGAAGCATGCAACAGGAAATAGCGCATGCTTCTAC 7560
Qy 7561 CATACAGCTCTTAAACAAAGGCTTAAAGGCAATTTCAAGTGAATTAACAGCTGAGAT 7620
Db 7561 CATACAGCTCTTAAACAAAGGCTTAAAGGCAATTTCAAGTGAATTAACAGCTGAGAT 7620
Qy 7621 AACGCTATATCTGACCAAGTTTGAAGCGAAATTCAGAGAGAAAGAAACAAACAAAGA 7680
Db 7621 AACGCTATATCTGACCAAGTTTGAAGCGAAATTCAGAGAGAAAGAAACAAACAAAGA 7680
Qy 7681 ACTATGCGCGAGATGAGACCAACGCGCCCTAGTGTATGATTAAGAGAAAGAAATGTC 7740
Db 7681 ACTATGCGCGAGATGAGACCAACGCGCCCTAGTGTATGATTAAGAGAAAGAAATGTC 7740
Qy 7741 TAGACTTTTAGGCAATCTGAAACCTGTTCAAGATGAATTCAGAAACTCCAAATCAAGAT 7800
Db 7741 TAGACTTTTAGGCAATCTGAAACCTGTTCAAGATGAATTCAGAAACTCCAAATCAAGAT 7800
Qy 7801 CAAAATCTTGAATAATTAACACTGAACCTTGGTCAAGATGATGCCATGCAAAAGTGA 7860
Db 7801 CAAAATCTTGAATAATTAACACTGAACCTTGGTCAAGATGATGCCATGCAAAAGTGA 7860
Qy 7861 AGTTGCAATTTTGCAGACCAAACTACTAAGCAGAAAGCGGAAAGCTGAGCTAAAGCGAT 7920
Db 7861 AGTTGCAATTTTGCAGACCAAACTACTAAGCAGAAAGCGGAAAGCTGAGCTAAAGCGAT 7920
Qy 7921 GCAGGTGAACCTTAACCTTAAAGCAAGATTAATCTTCAAGCTGCAATGAAGCAATAGAAA 7980
Db 7921 GCAGGTGAACCTTAACCTTAAAGCAAGATTAATCTTCAAGCTGCAATGAAGCAATAGAAA 7980
Qy 7981 CCTACAGAAATGCTTCCCAAGAGTGCAGTACATTAAGAGAAATTTGCAACCTTAA 8040
Db 7981 CCTACAGAAATGCTTCCCAAGAGTGCAGTACATTAAGAGAAATTTGCAACCTTAA 8040
Qy 8041 AACTAAGTGTGTAAGATTAAGAAATGAAATTAATTAATCTCAAAACCAACAGCCAGA 8100
Db 8041 AACTAAGTGTGTAAGATTAAGAAATGAAATTAATTAATCTCAAAACCAACAGCCAGA 8100
Qy 8101 GATTGCTTAAAGTCTTGTGTTGAAGATTAAGGAAGAAAGGCTTGGCTAGTTTAAAGA 8160
Db 8101 GATTGCTTAAAGTCTTGTGTTGAAGATTAAGGAAGAAAGGCTTGGCTAGTTTAAAGA 8160
Qy 8161 GGAACCTTAGGCGAGACAGCAGACAAAGATPACAAAGTTTGTCTTCAAAAGATTAATCA 8220
Db 8161 GGAACCTTAGGCGAGACAGCAGACAAAGATPACAAAGTTTGTCTTCAAAAGATTAATCA 8220
Qy 8221 GAAAGCTTCAACTTTCCCTGTAAGTGTGTTGGAAGTGTGTTGGAAGGAGATTAAGTGAAGCAGC 8280
Db 8221 GAAAGCTTCAACTTTCCCTGTAAGTGTGTTGGAAGTGTGTTGGAAGGAGATTAAGTGAAGCAGC 8280
Qy 8281 AATGCTTGTGCTGAGTGCAGAAAGAGCGGCTTGAAGAGAGAGCTGTACATTAATAGAA 8340
Db 8281 AATGCTTGTGCTGAGTGCAGAAAGAGCGGCTTGAAGAGAGAGAGCTGTACATTAATAGAA 8340
```



```

QY 8341 GAAATATCATCTTATATCAGCGACTATGTCAGAGTTCTGAAGATGCAAGAAACAAAGCC 8400
    |||||||
Db 8341 GAAATATCATCTTATATCAGCGACTATGTCAGAGTTCTGAAGATGCAAGAAACAAAGCC 8400
QY 8401 AAAATCTGATGCTCTTCTTCATCTGATCATCACAGAGGCTGCTCCACAAAC 8460
    |||||||
Db 8401 AAAATCTGATGCTCTTCTTCATCTGATCATCACAGAGGCTGCTCCACAAAC 8460
QY 8461 TGAACCTACAGACATGAGCGCTGTTACTCCAGAAAGGCTCTAAATGCCAGCTTACCT 8520
    |||||||
Db 8461 TGAACCTACAGACATGAGCGCTGTTACTCCAGAAAGGCTCTAAATGCCAGCTTACCT 8520
QY 8521 AGATCTCCAAAGAAAGTGGAGTCCAGACTAAACGTTTGTGTCACCAACAGCTGCGA 8580
    |||||||
Db 8521 AGATCTCCAAAGAAAGTGGAGTCCAGACTAAACGTTTGTGTCACCAACAGCTGCGA 8580
QY 8581 AATCTACAGCCAAATTATATGTCTCCAGGCAAGACCGGATGCAATATATCTTTC 8640
    |||||||
Db 8581 AATCTACAGCCAAATTATATGTCTCCAGGCAAGACCGGATGCAATATATCTTTC 8640
QY 8641 TCCAGCAAGGTTGAGCTGCACAAAGGCGCTGTCACCAACAGATCGAGATGCC 8700
    |||||||
Db 8641 TCCAGCAAGGTTGAGCTGCACAAAGGCGCTGTCACCAACAGATCGAGATGCC 8700
QY 8701 CACCCAGATGTCATATCCCTGGCAAGACCGGACTGCATAAATCTACTGAAAGCAC 8760
    |||||||
Db 8701 CACCCAGATGTCATATCCCTGGCAAGACCGGACTGCATAAATCTACTGAAAGCAC 8760
QY 8761 GTTATTGACAAATTGTCCTCCATGCAACAGCAAAAGTACAGAAATCTAAATTC 8820
    |||||||
Db 8761 GTTATTGACAAATTGTCCTCCATGCAACAGCAAAAGTACAGAAATCTAAATTC 8820
QY 8821 CCCTAAAGGCAAAATTATGATGTAAGTCAAAATGATGCTTACTGTCATCTAGTT 8880
    |||||||
Db 8821 CCCTAAAGGCAAAATTATGATGTAAGTCAAAATGATGCTTACTGTCATCTAGTT 8880
QY 8881 TTTTGAATATCTTAGTGTGATTTTTCAGACTCAACACAGAGCAATGACAA 8940
    |||||||
Db 8881 TTTTGAATATCTTAGTGTGATTTTTCAGACTCAACACAGAGCAATGACAA 8940
QY 8941 AAGTCAGGCTGAGAACCTGGTGTATGAGCAAAAGAAAGGCGACCTGATGTAAAC 9000
    |||||||
Db 8941 AAGTCAGGCTGAGAACCTGGTGTATGAGCAAAAGAAAGGCGACCTGATGTAAAC 9000
QY 9001 ATCTTAGATCCCTGTACATCTGACTCTGCTGTCCTCAAGAGACTGCTACTGCTT 9060
    |||||||
Db 9001 ATCTTAGATCCCTGTACATCTGACTCTGCTGTCCTCAAGAGACTGCTACTGCTT 9060
QY 9061 CTGTGAGAGAAACACTGAAACCTGCATGCTGATTAAGAGAGTCTGAGAAACAA 9120
    |||||||
Db 9061 CTGTGAGAGAAACACTGAAACCTGCATGCTGATTAAGAGAGTCTGAGAAACAA 9120
QY 9121 AAGTGTCTTTAGTAGTAATCACTGTTGGGCGAGTGGTTAGTCTTAAATTAAGTG 9180
    |||||||
Db 9121 AAGTGTCTTTAGTAGTAATCACTGTTGGGCGAGTGGTTAGTCTTAAATTAAGTG 9180
QY 9181 CAATAGCTCTCAGCTCTATTTATATGTTATGTTCTGTATTTGTTAGACTTTTAACTCC 9240
    |||||||
Db 9181 CAATAGCTCTCAGCTCTATTTATATGTTATGTTCTGTATTTGTTAGACTTTTAACTCC 9240
QY 9241 CTGTGACTCATATTTGGGCTCATCTAGTCTTGTGTGTGTGCGACACACACACTGCTG 9300
    |||||||
Db 9241 CTGTGACTCATATTTGGGCTCATCTAGTCTTGTGTGTGTGCGACACACACTGCTG 9300
QY 9301 AAAAATGACATTTGCAAGTATGTTTACTGACTGCTCTGCTGGGCGCAATCCATGCA 9360
    |||||||
Db 9301 AAAAATGACATTTGCAAGTATGTTTACTGACTGCTCTGCTGGGCGCAATCCATGCA 9360
QY 9361 AGCACCATTAGTGTGCGCAATGTTTTCACACTATATTAATGTCGACTTGTGAATA 9420
    |||||||
Db 9361 AGCACCATTAGTGTGCGCAATGTTTTCACACTATATTAATGTCGACTTGTGAATA 9420

```

```

QY 9421 GAAATATCTACAAAGATAGGCAAAAGTGTGTTTCTCACTAGTAAACGATATGGTTT 9480
    |||||||
Db 9421 GAAATATCTACAAAGATAGGCAAAAGTGTGTTTCTCACTAGTAAACGATATGGTTT 9480
QY 9481 AACTTAATTTCTCCTAATTAATAATTTCTATATATGCACTCTATTAGTGACCTAAT 9540
    |||||||
Db 9481 AACTTAATTTCTCCTAATTAATAATTTCTATATATGCACTCTATTAGTGACCTAAT 9540
QY 9541 GGTATTCTGCACTCTTGTAAACCTTTTAAATTCAGTCTAATTAAGACAGCCCTGAT 9600
    |||||||
Db 9541 GGTATTCTGCACTCTTGTAAACCTTTTAAATTCAGTCTAATTAAGACAGCCCTGAT 9600
QY 9601 TTTAAATTTAAAAAAAACGGAATTC 9626
    |||||||
Db 9601 TTTAAATTTAAAAAAAACGGAATTC 9626

```

RESULT 2

AA12953
ID AA12953 standard; DNA; 7684 BP.

XX AA12953;

DT 19-MAR-1999 (first entry)

DE Enterococcus faecalis genome contig SEQ ID NO:16.

KW Enterococcus faecalis; contig; detection; Enterococcal infection;

KW vaccine; attenuation; computer readable medium; ds.

OS Enterococcus faecalis.

PN W09850555-A2.

PD 12-NOV-1998.

PF 04-MAY-1998; 98WO-US08985.

PR 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

PR 16-MAY-1997; 97US-0046555.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Dillon PJ, Kunsch CA;

DR WPI: 1999-045171/04.

PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.

PS Claim 1; Page 316-320; 2084pp; English.

XX A computer readable medium has been developed which has recorded on it
XX 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
XX AA12938 to AA13919 represent these nucleotide sequences which are
XX primary nucleotide sequences, also known as contigs. The computer-based
XX system can identify fragments of the Enterococcus faecalis genome with
XX commercial importance. The products can be used to detect the presence
XX of Enterococcus faecalis in samples. They can also be used for
XX diagnosing Enterococcal infection in an animal and monitoring
XX progression of disease, and for identifying agents which can be used to
XX modulate the growth or pathogenicity of Enterococcus faecalis, or
XX another related organism, in vivo or in vitro. In particular the
XX polypeptides encoded by the Enterococcus faecalis nucleotide sequences
XX can be used in vaccines to prevent or attenuate an Enterococcal
XX infection.

SQ Sequence 7684 BP; 2518 A; 1290 C; 1674 G; 2191 T; 11 other;

Query Match 0.2%; Score 23; DB 20; Length 7684;

Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 23: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3973 AATTGAACACAGATGAATTAA 3995
|||||
Db 4287 AATTGAACACAGATGAATTAA 4309

RESULT 3

ABL08140
ID ABL08140 standard; cDNA; 15101 BP.

AC ABL08140;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 18902.

KW Drosophila: developmental biology; cell signalling; insecticide;

KM pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PE 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX MPI: 2001-6556860/75.

DR P-PSDB: ABB64037.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

PS Claim 1; SEQ ID NO 18902; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIGO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 15101 BP; 4366 A; 3330 C; 3468 G; 3937 T; 0 other;

Query Match 0.2%; Score 22; DB 23; Length 15101;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 22: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8784 CATGCAACACGCAAAAGTACA 8805
|||||
Db 13673 CATGCAACACGCAAAAGTACA 13694

RESULT 4

AAAX20600/c
ID AAAX20600 standard; DNA; 22191 BP.

AC AAAX20600;

XX 05-MAY-1999 (first entry)

DE Polynucleotide sequence from the genome of Treponema pallidum.

KW Treponema pallidum infection; syphilis; Borrelia infection; animal;

KM enzyme production; ds.

XX Treponema pallidum.

OS Treponema pallidum.

PN WO9859034-A2.

PD 30-DEC-1998.

PE 23-JUN-1998; 98WO-US13041.

PR 24-JUN-1997; 97US-0050667.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Fraser CM;

DR MPI: 1999-081273/07.

PT New isolated Treponema pallidum nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of T. pallidum infections, particularly syphilis

PS Claim 1; Page 673-685; 1150pp; English.

XX AAAX20500-21243 represent polynucleotide sequences from the genome of
CC Treponema pallidum. The sequences can be used for detection,
CC diagnosis, characterisation, prevention and therapy for T. pallidum
CC infections, particularly syphilis. They can also be used for detecting
CC diseases related to Borrelia infections in animals, and for the
CC production of biosynthetic products such as enzymes.

XX Sequence 22191 BP; 4744 A; 5303 C; 6215 G; 5916 T; 13 other;

Query Match 0.2%; Score 22; DB 20; Length 22191;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 22: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5419 AAAGATAAGAGTCAGCGCTG 5440
|||||
Db 12662 AAAGATAAGAGTCAGCGCTG 12641

RESULT 5

AAAX26825/c
ID AAAX26825 standard; DNA; 21 BP.

AC AAAX26825;

DT 22-JUN-1999 (first entry)

XX PCR primer for centromere-associated protein-E (CENP-E) DNA.

DE CENP-E; centromere-associated protein-E; ATPase activity;
KW plus end-directed microtubule motor activity; chromosome congression;
KW microtubule binding activity; chromosome movement; mitosis; PCR primer;
KW cell proliferation; tumor; metastasis; vascular malfunction;
KW inflammatory disease; immune disease; angiogenesis; hypertension;
KW restenosis; fungal infection; selective herbicide; fungicide;
KW insecticide; plant growth regulator; activator; cancer cell marker; ss.

XX Synthetic.

OS Xenopus sp.

PN WO9913061-A1.

PD 18-MAR-1999.

sequence 692 BP; 201 A; 150 C; 205 G; 136 T; 0 other;

QY 1830 ACTCAAGGAGGAGAGAAAAA 1850

РЕСУРС 7

AC ABL26462;
XY

Drosophila melanogaster genomic polynucleotide SEQ ID NO 30859

Drosophila mela

PD 27-SEP-2001.
XX

23-MAR-2000; 2000US-191637P.
11-JUL-2000; 2000US-0614150

Venter JC, Adams M, Li PWD, Myers EW;

New isolated nucleic acid detection reagent for detecting 1000 or more genes from *Drosophila* and *Escherichia coli*.

The invention relates to an isolated nucleic acid detection capable of detecting 1000 or more

useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB101840-AB130511), expressed DNA sequences (AB101840-AB116175) and the encoded proteins (AB57737-AB572072).

RESULT 10

AAS74043

ID AAS74043 standard; cDNA; 3054 BP.

AC AAS74043;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #9847.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KM food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PA Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR P-PSDB; ABG098356.

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity

XX Claim 1; SEQ ID NO 9847; 103bp; English.

PS The invention relates to isolated polynucleotide (I) and

XX polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. AAS64197-AAS94564 represent novel human

CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 3054 BP; 967 A; 705 C; 712 G; 670 T; 0 other;

SQ Query Match 0.2%; Score 21; DB 23; Length 3054;

DB Best Local Similarity 100.0%; Pred No. 51;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7478 TTGGAGGAATTGGAGATCTG 7498

DB 992 TTGAGGCAATTGGAGATCTG 1012

RESULT 11

ABL27156/c

ID ABL27156 standard; DNA; 4117 BP.

AC ABL27156;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 33941.

KW Drosophila; developmental biology; cell signalling; insecticide;

KM pharmaceutical; gene; ds.

XX Drosophila melanogaster.

OS WO200171042-A2.

PN 27-SEP-2001.

PD 23-MAR-2001; 2001WO-US09231.

PF 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

PA Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions

PT Claim 1; SEQ ID NO 33941; 21bp + Sequence Listing; English.

PS The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL1840-ABL16175) and the encoded proteins

CC (ABBS737-ABBS72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 4117 BP; 1190 A; 945 C; 975 G; 1007 T; 0 other;

Query Match 0.2%; Score 21; DB 23; Length 4117;

DB Best Local Similarity 100.0%; Pred. No. 52;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9598 GATTTTAATTAATAAAAAA 9618

DB 704 GATTTTAATTAATAAAAAA 684

RESULT 12

ABL12872/c

ID ABL12872 standard; cDNA; 4699 BP;

AC ABL12872;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 33098.

KW Drosophila; developmental biology; cell signalling; insecticide;

KM pharmaceutical; gene; ss.

XX Drosophila melanogaster.

OS WO200171042-A2.

XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI: 2001-656860/75.
XX P-PSDB; ABB68769.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1: SEQ ID NO 33098; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (AB57737-AB872072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 4699 BP; 1342 A; 1013 C; 1072 G; 1272 T; 0 other:
SQ
Query Match 0.2%; Score 21; DB 23; Length 4699;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 9598 GATTTAAATTAATAAAAAA 9618
|||||
DB 3862 GATTTAAATTAATAAAAAA 3842
RESULT 13
AAST73907/C
ID AAST73907 standard; cDNA; 7545 BP.
XX
XX AAST73907;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #9711.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
XX

DR P-PSDB; ABB09720.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
XX Claim 1: SEQ ID NO 9711; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging or sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 7545 BP; 1600 A; 1993 C; 1792 G; 2160 T; 0 other:
SQ
Query Match 0.2%; Score 21; DB 23; Length 7545;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 7478 TTGAGAGCAATTGAGACTCTG 7498
|||||
DB 6339 TTGAGAGCAATTGAGACTCTG 6319
RESULT 14
AAK74473
ID AAK74473 standard; DNA; 7971 BP.
XX
XX AAK74473;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:29285.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
XX
XX 04-FEB-2000; 2000US-0180628.
XX
XX 24-FEB-2000; 2000US-0184664.
XX
XX 02-MAR-2000; 2000US-0186350.
XX
XX 16-MAR-2000; 2000US-0189874.
XX
XX 17-MAR-2000; 2000US-0190076.
XX
XX 18-APR-2000; 2000US-0198123.
XX
XX 19-MAY-2000; 2000US-0205515.
XX
XX 07-JUN-2000; 2000US-0209467.
XX
XX 28-JUN-2000; 2000US-0214886.
XX

CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 7971 BP; 2011 A; 1701 C; 1666 G; 2593 T; 0 other;

Query Match 0.2%; Score 21; DB 22; Length 7971;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 794 ACAAAATTCAGAGAACTGTGAT 814
|||||
DB 2703 ACAAAATTCAGAGAACTGTGAT 2723

RESULT 15
AAK52052
ID AAK52052 standard; cDNA; 8617 BP.
XX
AC AAK52052;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 597.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KM tissue growth factor; immunomodulatory; cancer; leukaemia;
KM nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0634936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI: 2001-476283/51.
DR P-PSDB: AAM78919.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX
PS Claim 1; Page 2144-2154; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or

CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 8617 BP; 2468 A; 2025 C; 2201 G; 1923 T; 0 other;

Query Match 0.2%; Score 21; DB 22; Length 8617;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7478 TTGAGGAAATTTGAGATCTG 7498
|||||
DB 1207 TTGAGGAAATTTGAGATCTG 1227

Search completed: October 7, 2002, 11:17:13
Job time : 1337 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 7, 2002, 09:33:18 : Search time 11207 Seconds
(without alignments)
17974.377 Million cell updates/sec

Title: US-09-150-867-2

Perfect score: 9626
Sequence: 1 gaattcgcggatcgatagc.....ttaaaaaaaaaaacggaattc 9626

Scoring table: ~~ORF6-NUC~~
Gapop-60.0, Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size: 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

GenEmbl:
1: gb_ba:*
2: gb_hng:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hng_hum:*
31: em_hng_inv:*
32: em_hng_other:*
33: em_hng_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Match	Length	ID	Description
------------	-------	-------	--------	----	-------------

1	9610	99.8	9610	5	AF027728	AF027728 Xenopus 1
2	24	0.2	431	1	AB011407	AB011407 Buchnera
3	24	0.2	495	9	HUMTCBXN	M13575 Human T-cell
4	24	0.2	500	9	HUMTCBZE	M31347 Human T-cell
5	24	0.2	616	9	HSDNATCRA	X75418 H.sapiens T
6	24	0.2	737	9	HSTRVB132	X61445 H.sapiens T
7	24	0.2	8045	9	HSU07976	U07976 Human T cel
8	24	0.2	33037	9	HSU07978	U07978 Human T cel
9	24	0.2	72007	2	AC101336	AC101336 Mus muscu
10	24	0.2	135021	2	AC022170	AC022170 Homo sapi
11	24	0.2	215977	2	AC073415	AC073415 Homo sapi
12	24	0.2	267156	6	AX336388	AX336388 Sequence
13	24	0.2	267156	9	U66059	U66059 Human germ
14	24	0.2	293181	1	AP001119	AP001119 Buchnera
15	23	0.2	58777	3	AC006733	AC006733 Caenorhab
16	23	0.2	134470	2	AC094452	AC094452 Rattus no
17	23	0.2	153192	9	HSZA234K24	AL121895 Human DNA
18	23	0.2	169827	2	AC105732	AC105732 Oryza sat
19	23	0.2	177913	2	AC094889	AC094889 Rattus no
20	22	0.2	1397	1	AF302885	AF302885 Orientia
21	22	0.2	1397	1	AF302890	AF302890 Orientia
22	22	0.2	1397	1	AF302994	AF302994 Orientia
23	22	0.2	1422	1	AF173034	AF173034 Orientia
24	22	0.2	1424	1	AF173033	AF173033 Orientia
25	22	0.2	1424	1	AF173036	AF173036 Orientia
26	22	0.2	1426	1	AF173035	AF173035 Orientia
27	22	0.2	1544	1	AF050669	AF050669 Orientia
28	22	0.2	1551	1	RTU19903	U19903 Rickettsia
29	22	0.2	1819	10	MMU243485	AJ243485 Mus Muscu
30	22	0.2	4816	8	SCYBL047C	Z35808 S.cerevisia
31	22	0.2	11101	1	AE001191	AE001191 Treponema
32	22	0.2	22469	8	SC2KBF	X78214 S.cerevisia
33	22	0.2	32560	8	SC11L260	Z23261 S.cerevisia
34	22	0.2	33793	2	AC107550	AC107550 Rattus no
35	22	0.2	36998	3	CELC54A12	U28733 Caenorhabd
36	22	0.2	38523	3	CEY39A1B	AL021482 Caenorhab
37	22	0.2	38700	3	CET09A5	Z36753 Caenorhabd
38	22	0.2	49999	9	HSZ98266	Z98266 Homo sapien
39	22	0.2	51040	2	AC095816	AC095816 Rattus no
40	22	0.2	61156	2	AC097309	AC097309 Rattus no
41	22	0.2	68275	9	AC073335	AC073335 Homo sapi
42	22	0.2	72520	8	AC006570	AC006570 Arabidops
43	22	0.2	78079	2	AC015049	AC015049 Drosophi
44	22	0.2	83946	9	HSA123M24	AL078459 Human DNA
45	22	0.2	93443	8	ATF2206	AL050300 Arabidops

ALIGNMENTS

RESULT 1
AF027728 9610 bp mRNA linear VRT 15-NOV-1997
LOCUS Xenopus laevis kinesin-related protein (XCENP-E) mRNA, complete
DEFINITION cds
ACCESSION AF027728
VERSION AF027728.1 GI:2586070
KEYWORDS
SOURCE
ORGANISM
Xenopus laevis
African clawed frog.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE
1 (bases 1 to 9610)
Wood, R.W., Sakowicz, R., Goldstein, L.S. and Cleveland, D.W.
CENP-E is a plus end-directed kinetochore motor required for
metaphase chromosome alignment
JOURNAL
MEDLINE
98028574
2 (bases 1 to 9610)
Wood, R.W. and Cleveland, D.W.
REFERENCE
AUTHORS
TITLE
JOURNAL
Direct Submission
Submitted (01-OCT-1997) Cell Biology, Ludwig Institute for Cancer

Research, 9500 Gilman Dr., La Jolla, CA 92093-0660, USA
Location/Qualifiers
1..9610
/organism="Xenopus laevis"
/db_xref="taxon:8335"
1..9610
/gene="XCENP-E"
135..8999
/gene="XCENP-E"
/note="Kinetochore-associated; plus end-directed microtubule motor"
/codon_start=1
/product="Kinesin-related protein"
/protein_id="AAC60300.1"
/db_xref="GI:2586071"
/translation="MSBGDAVVCYRVPFLQREGDQANLQMKAGNNTISQVDGTS
FNEPRVFNHSTESQITQETIAPVPIIRSLQGYNGTIFAYGQTSQSGKTYTMAGTPNSLG
IIPQAIQEVFKIIDEIPNREFLARSMEIYNETVYKIDLCDDRKKPLELEIDENRNV
YVADLTELVMVPEHVIQWKKEKNRNGETKMHDSRSHTTFRMIVESRDRNTV
NSBNCDAVMVSHNLVDLAGSERASGTGAGVRLKGCNINSLFTLIGQYIKLSG
QAGGFINRDSKLTPIIQLNSLGNKATYITCTIPVSPDELTSLQFASAKHVRNP
HNVYLDEALLKRYREIILKQLEMLESSSEKQQAKEHTVLDEITKOLHE
REDRIWHLTJNVASQDSODQYKRRRTVAAGKIQNLHAGSVDDMLSRLEG
NRSKAKSDMPSPPEIDSVCTESFDDDLSDMDNGLDAENNLASXYTHREKSL
HOSMIDGQISDVQFHDSSKENQIYLPKDSGMACRASFKEKTSLOQOSKE
LEKELVQSFELKLALEEQLSYAKMLNEMVNRHESINAQVTDVEKHYRKEMV
LDGSGYVANSNDIODSSVDKSLSSSDECEIHEKMLEOKVLEEFTEINLKNKSEND
KQNSBQDFMESIQCEIYMAEKANALEIAMLNEMVNRHESINAQVTDVEKHYRKEMV
KENQTEFEELEKETQEHQAOLIHGELSKLYENAVENIINLEETIKREIADLKE
QEIQLAELRKADNLOKRVNRPDLVSQMSSEKICEITFQKQSLDAEVTYKDAKE
CSFLRNELKKEKEDTSNNYKQEKASLEKQLEKESNKKMADLOKLEQAF
NEIYVINGLAGKVPRLDSLEKRVSEKQLEALEKNALENEVYKLESEYFL
PNEVECLKNQISKAEEIMLKQEGHSAIISKOELIMOQOSQIOLDEVTHTOS
KVQQTEQYLEMKRMHDLFEKYTRNKSEADILREKENTMESVEYIAQTHHOL
EETIRDEKQLEKHYEYFOAQITFPTPLSDSLPSEKLYEVSQDPILENDYHNILA
LATERNNIIVCLTEKNSLKEOVIDLNTLOSLQOASIEKSDLOKPRODESEVYLL
LEMELKGLHLSQSLTEKIOLENTLEETLOLQLEKMTTTERNELQJNEDLAE
HDSLKQDLSNIRDSITQDEIRPAQSEIEKQOLDSFQOLDSQVSSSNHAY
ANDKSEILGYNSTOSQMLGEBDELQTSKALVSELELRAHVKSVEGENLETKL
NGELKEILGSESEVLSKMLENKEDNNKLEQAEYSERKQSLDEEYFSGSQLY
DELEVLQAOLKAAEERLEIKDRYFELVQANTNLVEGKLETPLOADHEDSDIRSE
EMELKIVGELERNQYLERLOEKLESLNKLLEKEMLETPLOADHEDSDIRSE
ENILTKENIDTTLKHSQDQAOLOKTOEOLAKNLAIASNDCPTORKEHSADCV
PLEEKILLTEELHOKTNEOEKILLHENELEQAOVLKQVHELMKSMIESKLSL
OHEKHDTQEOULLAKQOMOVYQEKKELOQTHHLRAVDHLENLELQVYNLNEAQ
KTKKEOCLLENKELEDSQHRQCEIIEBKSLKDESLLETLEKEQVYNLNEAQ
MVMHEMELKNSQRTVYERDQLODLRESVMSIETODLRAQAEALQOQKXQV
TSQISVLQEKISLENOMLYNATVETLSEBDLQOSQHLSEIETLSLKEEF
ALQAEKDKADARKTIDLEKISNIEEOLQOATNLKLETVRESLQCKQOLANT
EHLRETLKSDALGKMEQERDEAANKVIALTKMSLEKQLEINENVTTLKEBGEKET
FYLOPSQOOSQSELELRESLTKDLOLEAEKETSQATNETKNTTAKISLSEBITL
ONASILNEAVSERENLRHSKQOLVSELQSLTKRDRHAFQSKKEKEDINNLAEKV
AEIKITLTKMEDEFDSKESLQOSHSLELQVTELOMKQKQEDINNLAEKV
EVEDELLQSLLEQLOQLOMELRNKLENYELCEKMDIMEKISVLRQMLQNEP
DDVAENMDILIESNOEIOLEMEKISAVYSEHNTLSLSSELOKEIEARHOMLN
SLSSTLSRSQSLQLEOHGRKMSDASELEKCELEFVNTLAFKKNITQVODDEAYO
KRDEDLRLQDQLOEGRKMSDASELEKCELEFVNTLAFKKNITQVODDEAYO
VELNVOGSTLOELEHKQGMQLEERQGLHVDAKTSLKQSMOQBRRTASTQILLTK
LKAVQSKIORETTVYLNQFPAKLOEKQONKELRMRMHGFSASVMEENARLIGI
LTKVODESKLOSRIKMLENMLNVKQDMHGEKVAIILQDLKLSRAMEELNAAQV
LTKQDNLDPAKKEIENLQKVAQVAPKEIDNLTQVYITENEKIKEYATQOEL
ATUKSLEDEKBERLRLKELRAQADNTQVCPVQAGGAGTQVQAGGAGTQVQ
AMVLVQSEKAALERLSHYKRYKHLSPMSSEDEKRSKADSHSSGSHSGSP
HKETVYRHGPTPERSEMSLHIGSPKSESTKRVKVSNSREITVLSQMSGKGMH
KHLTSPKQGLKRLALSPNBSMPOHNTISPGKGLKHNLTLESTLFDNLSPCKQOK
VOENLNSPKGKLFQVYKSMYCPQSFQFNKSLGDFSELNLTAESNDKQAEAWMYEAK
KETAPCKTS"
misc_difference 1929..1935
/gene="XCENP-E"
/note="nucleotide sequence is absent in another XCENP-E
cDNA"

Query Match 99.8%; Score 9610; DB 5; Length 9610;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 9610; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 GAGTCGGATAGGCTTAGTCGGGCGAGGGAATTCAACTGGTTATAGAAACTTGAAC 68
DB 1 GAGTCGGATAGGCTTAGTCGGGCGAGGGAATTCAACTGGTTATAGAAACTTGAAC 60
QY GCCGCAAAAGGACCTAAGTAGCAGAGACAGGAGGGTGGCTGCCATTTCTCCAC 128
DB GCCGCAAAAGGACCTAAGTAGCAGAGACAGGAGGGTGGCTGCCATTTCTCCAC 120
QY 129 TAATCGGCTCTCAAAATCTCCGAGGAGATCAGTTAAAGTGTGTGAGGGTTCCGCC 188
DB 121 TAATCGGCTCTCAAAATCTCCGAGGAGATCAGTTAAAGTGTGTGAGGGTTCCGCC 180
QY 189 TTATACGAGAGAACAGGGGATCAACCACTGCATGGAATGGAAGGCTGGAACACACCA 248
DB 181 TTATACGAGAGAACAGGGGATCAACCACTGCATGGAATGGAAGGCTGGAACACACCA 240
QY 249 TTTCGCAAGTTGATGGGACAAAGCTCTTCAATTTGCATCGTGTATTTAATTCACGAAT 308
DB 241 TTTCGCAAGTTGATGGGACAAAGCTCTTCAATTTGCATCGTGTATTTAATTCACGAAT 300
QY 309 CAACAAGTCAAAATTTACCAAGAAATAGCACTACCTATATACATCAGATCAGCTTTCAGGAGT 368
DB 301 CAACAAGTCAAAATTTACCAAGAAATAGCACTACCTATATACATCAGATCAGCTTTCAGGAGT 360
QY 369 ATATAGGCAAAATTTGATGATGAGACAGACATCTTCAGCAAGACGTCACAAATGATG 428
DB 361 ATATAGGCAAAATTTGATGATGAGACAGACATCTTCAGCAAGACGTCACAAATGATG 420
QY 429 GAACACCAATTTATTTGGGATATATACCCCAAGCCATACAGAAAGTTTAAATTTATTC 488
DB 421 GAACACCAATTTATTTGGGATATATACCCCAAGCCATACAGAAAGTTTAAATTTATTC 480
QY 489 AGGAGATACGGAAGAGAGAGTTCTCTTAAGATTTCTTAATGAGATTTCAATGAAA 548
DB 481 AGGAGATACGGAAGAGAGAGTTCTCTTAAGATTTCTTAATGAGATTTCAATGAAA 540
QY 549 CTGTGAAAGACCTACCTGTGTGATGACAGAAAGAACCCCTTGGAAATTCGCGAGGAT 608
DB 541 CTGTGAAAGACCTACCTGTGTGATGACAGAAAGAACCCCTTGGAAATTCGCGAGGAT 600
QY 609 TTATAGAAACGCTATGTGTTGCTGACCTGACAGAACTGTATAGTTCCTGACATG 668
DB 601 TTATAGAAACGCTATGTGTTGCTGACCTGACAGAACTGTATAGTTCCTGACATG 660
QY 669 TAATACAGTGTATCAAAAAGGTGAAAAAACAGACATTTATGAGAGACTAAATGATG 728
DB 661 TAATACAGTGTATCAAAAAGGTGAAAAAACAGACATTTATGAGAGACTAAATGATG 720
QY 729 ATCATAGTGTGCTTACATTAATATATAGATATGTTGAAGCCGACAGACGAATG 788
DB 721 ATCATAGTGTGCTTACATTAATATATAGATATGTTGAAGCCGACAGACGAATG 780
QY 789 ATCCCAAAATTCAGAGAACTGTGATGAGAGTGTATCTCACTTGAATTTGGTAG 848
DB 781 ATCCCAAAATTCAGAGAACTGTGATGAGAGTGTATCTCACTTGAATTTGGTAG 840
QY 849 ACTTGTGCTGAGTAAAGAGACAGCCAACTGAGCTGGAAGGTGTGAGACTTAAGAG 908
DB 841 ACTTGTGCTGAGTAAAGAGACAGCCAACTGAGCTGGAAGGTGTGAGACTTAAGAG 900
QY 909 GGTGCAACATCAACGCGAGCTGTATCTTGGACAGGTATTAAGAGGTTTGGCAG 968
DB 901 GGTGCAACATCAACGCGAGCTGTATCTTGGACAGGTATTAAGAGGTTTGGCAG 960
QY 969 GCCAGGCTGTGTGATTTATTAACATACAGACAGCAAACTCACCAGAAATTCCTCAAAAT 1028
DB 961 GCCAGGCTGTGTGATTTATTAACATACAGACAGCAAACTCACCAGAAATTCCTCAAAAT 1020

```
QY 1029 CATGGGAGAAATGCTAAAACGGTTATATTGCAATATACCCAGTTCTTTGATG 1088
|||||
Db 1021 CATGGGAGAAATGCTAAAACGGTTATATTGCAATATACCCAGTTCTTTGATG 1080
QY 1089 AGACTAAGTACACTTCAGTTGGCCAGTACTGCCAAACATGAGAAATACCCCATG 1148
|||||
Db 1081 AGACTAAGTACACTTCAGTTGGCCAGTACTGCCAAACATGAGAAATACCCCATG 1140
QY 1149 TTAATGAGTCTCGATGATGAAAGCGTTGCTAAAAGGTACAGAAAGAAATCTTGAT 1208
|||||
Db 1141 TTAATGAGTCTCGATGATGAAAGCGTTGCTAAAAGGTACAGAAAGAAATCTTGAT 1200
QY 1209 TAAAGAAACATTTAGAAATTTAGAGTCAATCGTCTGAACAAAAGCTCAAGCATGGCT 1268
|||||
Db 1201 TAAAGAAACATTTAGAAATTTAGAGTCAATCGTCTGAACAAAAGCTCAAGCATGGCT 1260
QY 1269 AAGAAGAGCATACAGTGTGCTAGCTGAATCAACACATACCAAGAGAGAGAT 1328
|||||
Db 1261 AAGAAGAGCATACAGTGTGCTAGCTGAATCAACACATACCAAGAGAGAGAT 1320
QY 1329 GAATATGCACTTGACAAATATTTGTTGCTTCATCCCAAGAAATCTCAACAGACCATA 1388
|||||
Db 1321 GAATATGCACTTGACAAATATTTGTTGCTTCATCCCAAGAAATCTCAACAGACCATA 1380
QY 1389 GGGTCAACGAAACGAAAGATTACGTGGGCCGACGAAAAATCCAAATAGTTTACATG 1448
|||||
Db 1381 GGGTCAACGAAACGAAAGATTACGTGGGCCGACGAAAAATCCAAATAGTTTACATG 1440
QY 1449 CTTCGGTCTTCTGACTTGTATGCTATCCGATTTACCTGGCATTTTACAGAGAG 1508
|||||
Db 1441 CTTCGGTCTTCTGACTTGTATGCTATCCGATTTACCTGGCATTTTACAGAGAG 1500
QY 1509 CAAAGTCTCTGACATGCTTCATTTCCAGAAATTTGATGATGATCTGTTTGTACAGAT 1568
|||||
Db 1501 CAAAGTCTCTGACATGCTTCATTTCCAGAAATTTGATGATGATCTGTTTGTACAGAT 1560
QY 1569 CTGATTTTATGACGCCCTCTCATGATGACAGCAATGGAATAGATGCAAGATGGAAT 1628
|||||
Db 1561 CTGATTTTATGACGCCCTCTCATGATGACAGCAATGGAATAGATGCAAGATGGAAT 1620
QY 1629 TAGCCAGTAAATTAATCTCAGAGAAAGAAACATCACTTCATCAATATGATGAGCTTG 1688
|||||
Db 1621 TAGCCAGTAAATTAATCTCAGAGAAAGAAACATCACTTCATCAATATGATGAGCTTG 1680
QY 1689 GACAGATTTCTGACAGTGTTCAGTTTCATGATTTCTTAAGAAAAACGACTACAAAT 1748
|||||
Db 1681 GACAGATTTCTGACAGTGTTCAGTTTCATGATTTCTTAAGAAAAACGACTACAAAT 1740
QY 1749 TCCCAAAAGACTCTGGTGTATGCGTGAATGCGAAAAAGCTTCTTTTGAAGAAAGATCA 1808
|||||
Db 1741 TCCCAAAAGACTCTGGTGTATGCGTGAATGCGAAAAAGCTTCTTTTGAAGAAAGATCA 1800
QY 1809 CAAGCTCCAGAACAACTACAGTCAAGAGAGAGAAAAAGAACTTGTAAGAACT 1868
|||||
Db 1801 CAAGCTCCAGAACAACTACAGTCAAGAGAGAGAAAAAGAACTTGTAAGAACT 1860
QY 1869 TCGAGCTCAAGATAGCAGAACTGGAAGACAGCACTAGTGTCAAAAGCTAAAAATCTAGAGA 1928
|||||
Db 1861 TCGAGCTCAAGATAGCAGAACTGGAAGAGACAGCACTAGTGTCAAAAGCTAAAAATCTAGAGA 1920
QY 1929 TGGTTACAAACTCGAGAGAGCATTTCCATTATGCTGAAGTCCAAACAGATGTTGAAGAG 1988
|||||
Db 1921 TGGTTACAAACTCGAGAGAGCATTTCCATTATGCTGAAGTCCAAACAGATGTTGAAGAG 1980
QY 1989 AAGTTGTGAAAAAGAAATGTCAGTCTTGAGAGACTGTGTTACATGATCAAAAGCT 2048
|||||
Db 1981 AAGTTGTGAAAAAGAAATGTCAGTCTTGAGAGACTGTGTTACATGATCAAAAGCT 2040
QY 2049 ACCTACAGATAGTTCTGTTGATGATGTAAGAGCTCTTAAGAGCTCCATGATGATGATAG 2108
|||||
Db 2041 ACCTACAGATAGTTCTGTTGATGATGTAAGAGCTCTTAAGAGCTCCATGATGATGATAG 2100
QY 2109 AACACGAAAAATGCTGGAACAAAGATGCTTATTTAGAGAGTTTATTTGAAGAACTTA 2168
|||||
Db 2101 AACACGAAAAATGCTGGAACAAAGATGCTTATTTAGAGAGTTTATTTGAAGAACTTA 2160
QY 2169 ACAAGAAAAAGTGAATGATTAACAAAAATCTTCTGAGCAAGATTTTATGAGAGATATTC 2228
|||||
Db 2161 ACAAGAAAAAGTGAATGATTAACAAAAATCTTCTGAGCAAGATTTTATGAGAGATATTC 2220
QY 2229 AGCTATGTGAAGCTATTAATGCGAGAAAAAGCAATATGCACTGAGAGAACTGGCACTTATGC 2288
|||||
Db 2221 AGCTATGTGAAGCTATTAATGCGAGAAAAAGCAATATGCACTGAGAGAACTGGCACTTATGC 2280
QY 2289 GAGATATTTTGAACAAATATTTCTAGAGAAATGAACTCTAAAAAGGAAATGCAATGC 2348
|||||
Db 2281 GAGATATTTTGAACAAATATTTCTAGAGAAATGAACTCTAAAAAGGAAATGCAATGC 2340
QY 2349 TGGAGCTTCACCTCAAGAGAAATCAAGAAACCAATGAGTTTGAATTTCTGAGAGAGAA 2408
|||||
Db 2341 TGGAGCTTCACCTCAAGAGAAATCAAGAAACCAATGAGTTTGAATTTCTGAGAGAGAA 2400
QY 2409 CTCAAAAAGAACACAGAGCAACACTAATCCATGAGATGCGAGTTTAAAGAAATGATG 2468
|||||
Db 2401 CTCAAAAAGAACACAGAGCAACACTAATCCATGAGATGCGAGTTTAAAGAAATGATG 2460
QY 2469 AAAATGAGAGATGTCAATCAAAAATCTGAGAAAGTCTAGAAACCTAAGAAACCTTC 2528
|||||
Db 2461 AAAATGAGAGATGTCAATCAAAAATCTGAGAAAGTCTAGAAACCTAAGAAACCTTC 2520
QY 2529 TGAAGAGCAGGAATTTCAACTTGCAAGAAATTAAGAAACCGCAGATTAACCTGCACAAA 2588
|||||
Db 2521 TGAAGAGCAGGAATTTCAACTTGCAAGAAATTAAGAAACCGCAGATTAACCTGCACAAA 2580
QY 2589 AAGTACGAAATTTTGAATCTCTCGGTTTCCATGGGTATAGTGAAGAACTGTGAGAGAA 2648
|||||
Db 2581 AAGTACGAAATTTTGAATCTCTCGGTTTCCATGGGTATAGTGAAGAACTGTGAGAGAA 2640
QY 2649 TCTTTCAACTGAAGCAATCTTTCTGATGCTGAGCTGAGCTCGGATGCTCAGAGAG 2708
|||||
Db 2641 TCTTTCAACTGAAGCAATCTTTCTGATGCTGAGCTGAGCTCGGATGCTCAGAGAG 2700
QY 2709 AATGTTCTTTCTCAGAAAGTGAAGAAATCTAGAGCTGAAGAGAAATGAGAGACATCA 2768
|||||
Db 2701 AATGTTCTTTCTCAGAAAGTGAAGAAATCTAGAGCTGAAGAGAAATGAGAGACATCA 2760
QY 2769 ACTGATTAATCAAAAAAGAGCTGCTGTTGTTGGAAGAGAGCTGGAAGAGTAA 2828
|||||
Db 2761 ACTGATTAATCAAAAAAGAGCTGCTGTTGTTGGAAGAGAGCTGGAAGAGTAA 2820
QY 2829 AATCAAACTCAAGAAAAATGGAAGCTGATTTGCGAAGAAAGTTGCAAAAGTCTTTTAA 2888
|||||
Db 2821 AATCAAACTCAAGAAAAATGGAAGCTGATTTGCGAAGAAAGTTGCAAAAGTCTTTTAA 2880
QY 2889 AGATTAACCTTAATAGGCTTCTGCGAGAGAAAGTCCCGAGAGATTTGCTTCTGCTG 2948
|||||
Db 2881 AGATTAACCTTAATAGGCTTCTGCGAGAGAAAGTCCCGAGAGATTTGCTTCTGCTG 2940
QY 2949 TTTGAATTTAGAAAAAGGTTTCTGAGTCTCAAAAGCAGCTTGAGAAAGCATTTGGAAGAA 3008
|||||
Db 2941 TTTGAATTTAGAAAAAGGTTTCTGAGTCTCAAAAGCAGCTTGAGAAAGCATTTGGAAGAA 3000
QY 3009 AAAATGCTTGGAGAAATGAAGTGAATGCTTATCAGAAATCAAAATTTTTCGCAAAATGAG 3068
|||||
Db 3001 AAAATGCTTGGAGAAATGAAGTGAATGCTTATCAGAAATCAAAATTTTTCGCAAAATGAG 3060
QY 3069 TTTGAATGCTTGAAGAAATCAGATCAGCAAGGCTTCTGAAGAGATTAATGTTTAAAGCAAG 3128
|||||
Db 3061 TTTGAATGCTTGAAGAAATCAGATCAGCAAGGCTTCTGAAGAGATTAATGTTTAAAGCAAG 3120
QY 3129 AAGGAGAACATTTCTGATCTATTTATAGCAAAACAGAGATTTATCATTCAGAGACAACTG 3188
|||||
Db 3121 AAGGAGAACATTTCTGATCTATTTATAGCAAAACAGAGATTTATCATTCAGAGACAACTG 3180
QY 3189 AGCAGATTTTACAACTGACTGAGAGAGTGAACACACACAGTCAAAAAGTCCAGACAGCTG 3248
|||||
```

Db 3181 AGCAGATTTTACAAGTGAACGAGTGAACACACACATCAAAAGATGACAGACTG 3240
QY 3249 AAGACATATCTTGAGATGAAGAAATGATGATCTTTTGAAGATATATACGA 3308
|||
Db 3241 AAGAGCAATCTTGAGATGAAGAAATGATGATCTTTTGAAGATATATACGA 3300
QY 3309 ACAAAGTGAAGCTGAAGACCTTTTAAGAGAAATGAGAACTTTAAAGGCATATGAGT 3368
|||
Db 3301 ACAAAGTGAAGCTGAAGACCTTTTAAGAGAAATGAGAACTTTAAAGGCATATGAGT 3360
QY 3369 CTGTGGAAGTAAAGATTTGTCAGACAAATATGAACTGTAAGAACTATTAAGGATTAAG 3428
|||
Db 3361 CTGTGGAAGTAAAGATTTGTCAGACAAATATGAACTGTAAGAACTATTAAGGATTAAG 3420
QY 3429 AGCAACTGCTTCATGAGAAAAATTTTTCAGCAATGACAGATATATTTCCGATTA 3488
|||
Db 3421 AGCAACTGCTTCATGAGAAAAATTTTTCAGCAATGACAGATATATTTCCGATTA 3480
QY 3489 CACCTCTTTAGACTGCTGCTTCCCTCAAAATTTAGTGAAGGAACTCTCAAGACCCA 3548
|||
Db 3481 CACCTCTTTAGACTGCTGCTTCCCTCAAAATTTAGTGAAGGAACTCTCAAGACCCA 3540
QY 3549 TAGAAATCAATGACTACCAATTTTATAGCCCTGCTCAGAAAGAAACAAATATTAAG 3608
|||
Db 3541 TAGAAATCAATGACTACCAATTTTATAGCCCTGCTCAGAAAGAAACAAATATTAAG 3600
QY 3609 TGTGCTAGAGACTGAAGAAACAGTCTCAGAGCAAGTATTTATTTGAACACTCAAC 3668
|||
Db 3601 TGTGCTAGAGACTGAAGAAACAGTCTCAGAGCAAGTATTTATTTGAACACTCAAC 3660
QY 3669 TTCAAAGTCTTCAAGCAACAAGATTTGAAGAAAGTCTATCTCCAGAAAGCCAAAGCAAGCT 3728
|||
Db 3661 TTCAAAGTCTTCAAGCAACAAGATTTGAAGAAAGTCTATCTCCAGAAAGCCAAAGCAAGCT 3720
QY 3729 TGGAAAGAGAGAGTTAAACTGTTTGGAGATGGAATCTAGTAAGGACACTAACTG 3788
|||
Db 3721 TGGAAAGAGAGAGTTAAACTGTTTGGAGATGGAATCTAGTAAGGACACTAACTG 3780
QY 3789 ACTCAGAGTCTCTATAGAGAAATTTGACAGTGAAGATTTGGAAGTTCAGAAAAACTGC 3848
|||
Db 3781 ACTCAGAGTCTCTATAGAGAAATTTGACAGTGAAGATTTGGAAGTTCAGAAAAACTGC 3840
QY 3849 AAACACTTCAGAGAGAGATGAAGAAATATTTACTATAGAAAGAAATGAGCTTCGACCAACT 3908
|||
Db 3841 AAACACTTCAGAGAGAGATGAAGAAATATTTACTATAGAAAGAAATGAGCTTCGACCAACT 3900
QY 3909 TCGAAGACTGTAAGCAGAGCATGATAGCTTAAACAAAGACCTTAGTGAACATTTGAGC 3968
|||
Db 3901 TCGAAGACTGTAAGCAGAGCATGATAGCTTAAACAAAGACCTTAGTGAACATTTGAGC 3960
QY 3969 ACTCAATTAAGAAACAAGATGAATTAAGGCTGCCAGAGAGAGCTTAAGAGAACAGAGC 4028
|||
Db 3961 ACTCAATTAAGAAACAAGATGAATTAAGGCTGCCAGAGAGAGCTTAAGAGAACAGAGC 4020
QY 4029 AACTGTTGATAGCTTTAGACAAACAGCTTTTATGATTTCTGTAGAGATTTTATCATACCA 4088
|||
Db 4021 AACTGTTGATAGCTTTTAGACAAACAGCTTTTATGATTTCTGTAGAGATTTTATCATACCA 4080
QY 4089 ACCATGATGCACTTCTTAACAGAGAAAGGTGCTATTGGTGAAGTTAATTCGTTACAA 4148
|||
Db 4081 ACCATGATGCACTTCTTAACAGAGAAAGGTGCTATTGGTGAAGTTAATTCGTTACAA 4140
QY 4149 GTGAAATGCTGCTGCTGAAGAGATGAGCTTCAAACTCTTTGTAAGCATTTAGTTTCAG 4208
|||
Db 4141 GTGAAATGCTGCTGCTGAAGAGATGAGCTTCAAACTCTTTGTAAGCATTTAGTTTCAG 4200
QY 4209 AACTGAGACTTCTGCTGCTGATGTAATAATCTGTGAGAGGAGAAACCTTAAATTCACAA 4268
|||
Db 4201 AACTGAGACTTCTGCTGCTGATGTAATAATCTGTGAGAGGAGAAACCTTGAATTCACAA 4260
QY 4269 AAAAATCTAATGCTTGAAGAAAGAGATATTGGCAATCTGAAGAAAGCAAGTGTGA 4328
|||
Db 4261 AAAAATCTAATGCTTGAAGAAAGAGATATTGGCAATCTGAAGAAAGCAAGTGTGA 4320

QY 4329 AGTCATGTTGGAGATCTTAAAGAAAGACACAAATTAAGCTCAAGACAGCAGAGAT 4388
|||
Db 4321 AGTCATGTTGGAGATCTTAAAGAAAGACACAAATTAAGCTCAAGACAGCAGAGAT 4380
QY 4389 ATTCTAGTAAAGAAATCAATTCAGTTTGAAGAGAGTGTGCTGCTCAGAGAGCTGG 4448
|||
Db 4381 ATTCTAGTAAAGAAATCAATTCAGTTTGAAGAGAGTGTGCTGCTCAGAGAGCTGG 4440
QY 4449 TAGATGATAGAGTCTGAAAGCAGCTTAAAGCAGCAGAAAGAGAGCTGAAATTA 4508
|||
Db 4441 TAGATGATAGAGTCTGAAAGCAGCTTAAAGCAGCAGAAAGAGAGCTGAAATTA 4500
QY 4509 AGGATGAGATTAATTTGAACCTTTACAACTGCAACCAATTTTACTTGAAGGAAAT 4568
|||
Db 4501 AGGATGAGATTAATTTGAACCTTTGTAACCTTTGTAACCTGCAACCAATTTTACTTGAAGGAAAT 4560
QY 4569 TGGAACTCCATTTGCAAGCTGACATGAGAGAGAGAGATGATGCGGCTTCTGAAGAA 4628
|||
Db 4561 TGGAACTCCATTTGCAAGCTGACATGAGAGAGAGAGATGATGCGGCTTCTGAAGAA 4620
QY 4629 TGGATATTAAGTCTTGGAGAAAACTTTGAGCAATCAGTATTTTACTGGAAGATTC 4688
|||
Db 4621 TGGATATTAAGTCTTGGAGAAAACTTTGAGCAATCAGTATTTTACTGGAAGATTC 4680
QY 4689 AAGAGAAAGCTGCACTGCTTAAACAACTTGAATCTTCAAAAGAGATGAGAGCT 4748
|||
Db 4681 AAGAGAAAGCTGCACTGCTTAAACAACTTGAATCTTCAAAAGAGATGAGAGCT 4740
QY 4749 CCGTCTTATTAAGATGACCTGCAACAGAGCTAGAAAGCTTCTGCTAGTGAACATTA 4808
|||
Db 4741 CCGTCTTATTAAGATGACCTGCAACAGAGCTAGAAAGCTTCTGCTAGTGAACATTA 4800
QY 4809 TTTTAAAGAGATTAATTAAGCAACCTTAAGCAATCTGATGATTAAGTCAAGCTAGCTGC 4868
|||
Db 4801 TTTTAAAGAGATTAATTAAGCAACCTTAAGCAATCTGATGATTAAGTCAAGCTAGCTGC 4860
QY 4869 AGAAAACAGCAGAGAGCTAGCTAGCTTGAAGATCTTGAATGCTGCTTGCACAT 4928
|||
Db 4861 AGAAAACAGCAGAGAGCTAGCTAGCTTGAAGATCTTGAATGCTGCTTGCACAT 4920
QY 4929 GTCCATTAACCTCAAGAAAAAGAAACCTCTGACATTTGTGATCTCTGGAAGAAAA 4988
|||
Db 4921 GTCCATTAACCTCAAGAAAAAGAAACCTCTGACATTTGTGATCTCTGGAAGAAAA 4980
QY 4989 TATGTTATTAACCTGAAGAAATTTGCATCAAAATTAATGAACAGGAAATTTCTCATG 5048
|||
Db 4981 TATGTTATTAACCTGAAGAAATTTGCATCAAAATTAATGAACAGGAAATTTCTCATG 5040
QY 5049 AAAAGAAATGAACCTTGAACAGCTCAGGTTAGCTTAAAGTGCAGAGGAGAAATCTGATGA 5108
|||
Db 5041 AAAAGAAATGAACCTTGAACAGCTCAGGTTAGCTTAAAGTGCAGAGGAGAAATCTGATGA 5100
QY 5109 AGATATGATGCAATGCAATCTCCTCACTTGAAGTCTTGAAGCTTGAAGCAATGATCTG 5168
|||
Db 5101 AGATATGATGCAATGCAATCTCCTCACTTGAAGTCTTGAAGCTTGAAGCAATGATCTG 5160
QY 5169 AACCAACACTTCTGCTTAAACAGAGATGAAGTGAAGTACTCAAGAAAGAAAGAGAGC 5228
|||
Db 5161 AACCAACACTTCTGCTTAAACAGAGATGAAGTGAAGTACTCAAGAAAGAAAGAGAGC 5220
QY 5229 TGCACAAACCCATGAACACTTAAACAGCTGAGGTGAGACCATCTAAGAAAGAAATTAAT 5288
|||
Db 5221 TGCACAAACCCATGAACACTTAAACAGCTGAGGTGAGACCATCTAAGAAAGAAATTAAT 5280
QY 5289 TGGGTTTGAATTTTAAAGATGAAGCCAGCAAAAGCCACTAAAGAGCAATGTCTGA 5348
|||
Db 5281 TGGGTTTGAATTTTAAAGATGAAGCCAGCAAAAGCCACTAAAGAGCAATGTCTGA 5340
QY 5349 ATGAGATTAAGAACTTGAAGCAAGCCAGACAGATCTTAAGTGAAGTGAAGAGCTTA 5408
|||
Db 5341 ATGAGATTAAGAACTTGAAGCAAGCCAGACAGATCTTAAGTGAAGTGAAGAGCTTA 5400

QY	5409	TGAAAAGCTTAAAGGATAAAGAGCTCAGCGCTGGAAACTTTAAAGAAATCTGACAAAAG	5468
Db	5401	TGAAAAGCTTAAAGGATAAAGAGCTCAGCGCTGGAAACTTTAAAGAAATCTGACAAAAG	5460
QY	5469	TAAATTAATCTAACCAAGAAATGGAATGGTTATGCTGGAATGGAGAAATTTGAAAAATA	5528
Db	5461	TAAATTAATCTAACCAAGAAATGGAATGGTTATGCTGGAATGGAGAAATTTGAAAAATA	5520
QY	5529	GCCAGAGACTGTAAATGTCTGAGAGGGAGCAGCTGCGAAGACACCTTAAGGAGAGAGTGTG	5588
Db	5521	GCCAGAGAGACTGTAAATGTCTGAGAGGGAGCAGCTGCGAAGACACCTTAAGGAGAGTGTG	5580
QY	5589	AAATGTCCATTGAAACTCAAGATGATCTTAAGAAAGGCTCAGAGACATTGCAGCAGAGA	5648
Db	5581	AAATGTCCATTGAAACTCAAGATGATCTTAAGAAAGGCTCAGAGACATTGCAGCAGAGA	5640
QY	5649	AAGTAAAGTTACAGGAACCTACCCTCCAGATTTCTGCTGAGGAAAAAGATCTCTCTC	5708
Db	5641	AAGTAAAGTTACAGGAACCTACCCTCCAGATTTCTGCTGAGGAAAAAGATCTCTCTC	5700
QY	5709	TGGAATAATCAGATGTCTTTATATATGTTCGCAACTGTGAAAGAAACTCTTAAGCCGAAAGAGATG	5768
Db	5701	TGGAATAATCAGATGTCTTTATATATGTTCGCAACTGTGAAAGAAACTCTTAAGCCGAAAGAGATG	5760
QY	5769	ACCTGAAACCACTCTAAGCAACACCTGTCTCAGAAATTTAAACTTTAGTCTGTCTTAA	5828
Db	5761	ACCTGAAACCACTCTAAGCAACACCTGTCTCAGAAATTTAAACTTTAGTCTGTCTTAA	5820
QY	5829	AAGAAAAAGATTTGCATTGTGSAACAAGCAGAGAGAGACAAACCTGATCTGCGCAGGAAA	5888
Db	5821	AAGAAAAAGATTTGCATTGTGSAACAAGCAGAGAGAGACAAACCTGATCTGCGCAGGAAA	5880
QY	5889	CAATAGATATCACAGAGAAAGATATCAATAATAGAAAGAACAGTTACTTCAACAGCCACA	5948
Db	5881	CAATAGATATCACAGAGAAAGATATCAATAATAGAAAGAACAGTTACTTCAACAGCCACA	5940
QY	5949	ATTTAAAGGAACTCTTTATGAAAGAGAGAGCTTATCTCAGTGTAAAGGACACTGGCTT	6008
Db	5941	ATTTAAAGGAACTCTTTATGAAAGAGAGAGCTTATCTCAGTGTAAAGGACACTGGCTT	6000
QY	6009	TGAACACAGAAACACTTAGGGAACAACTTGAAGAGCAACACTTGGCATTTGGGTAAATGG	6068
Db	6001	TGAACACAGAAACACTTAGGGAACAACTTGAAGAGCAACACTTGGCATTTGGGTAAATGG	6060
QY	6069	AGCAGAGAGAGATGGAAGCTGCCCCAATAAAGTAAATAGCTTTTCAGAAAAAGATGTCTCTC	6128
Db	6061	AGCAGAGAGAGATGGAAGCTGCCCCAATAAAGTAAATAGCTTTTCAGAAAAAGATGTCTCTC	6120
QY	6129	TAGAAGACACGATCAATGAAATGTTACTACTCTTAAGAGAGGTGAGGTTGAAAAAGAGA	6188
Db	6121	TAGAAGACACGATCAATGAAATGTTACTACTCTTAAGAGAGGTGAGGTTGAAAAAGAGA	6180
QY	6189	CCTTCTACCTTCAGAGACCTTCAACACACAGCTCTTCCGAGATGGAAGACAGCTCAGAG	6248
Db	6181	CCTTCTACCTTCAGAGACCTTCAACACACAGCTCTTCCGAGATGGAAGACAGCTCAGAG	6240
QY	6249	AGCTTTAAAGACCAAGATTTGCAGTTGTGGAAGAGGCGCAGAGAGAGATTAAGTAACTGA	6308
Db	6241	AGCTTTAAAGACCAAGATTTGCAGTTGTGGAAGAGGCGCAGAGAGAGATTAAGTAACTGA	6300
QY	6309	CTAATGAATAAAGAAATCTCACTGCTTAAGATCTCTTCTCTAGAGAGAGATTTCTTCAGA	6368
Db	6301	CTAATGAATAAAGAAATCTCACTGCTTAAGATCTCTTCTCTAGAGAGAGATTTCTTCAGA	6360
QY	6369	ATGCTACATTTTGAATGAAGCTGTAAAGGAAAGGAAAAACCTTGCCATTTCGAAGAGC	6428
Db	6361	ATGCTACATTTTGAATGAAGCTGTAAAGGAAAGGAAAAACCTTGCCATTTCGAAGAGC	6420
QY	6429	AACGTGTTTCAGAAATTTGAGACACACTCTCACTGACACTTAAAGAGTGTAGACCATGTTG	6488
Db	6421	AACGTGTTTCAGAAATTTGAGACACACTCTCACTGACACTTAAAGAGTGTAGACCATGTTG	6480
QY	6489	CGCAATCTAAACGCGAAAAAGATGAAGCTGTAAATTAATAGCCAGTCTCGCTGAGAAA	6548

[illegible]

QY	8709	ATGTCATATCCCTGGCAGACCGGACCTGCAATAAAATCTAACTGAAAGCACGTTATTCG	8768
Db	8701	ATGTCATATCCCTGGCAGACCGGACCTGCAATAAAATCTAACTGAAAGCACGTTATTCG	8760
QY	8769	ACATATTTGCTTTCGCATGCAAAACAGCAAAAAAGTACAGAAATCTAAATTTCCCTTAAG	8828
Db	8761	ACATATTTGCTTTCGCATGCAAAACAGCAAAAAAGTACAGAAATCTAAATTTCCCTTAAG	8820
QY	8829	GCAATTTATTTGATGTGAAGTCAAAATCGATGCCCTTACTGTCCATCTCAGTTTTTTGATA	8888
Db	8821	GCAAAATTAATTTGATGTGAAGTCAAAATCGATGCCCTTACTGTCCATCTCAGTTTTTTGATA	8880
QY	8889	ATTCTTAAGCTGGGTGATTTTTTTCAGAGCTCAACACAGCAGAGAGCAATGACAAAAGTCAG	8948
Db	8881	ATTCTTAAGCTGGGTGATTTTTTTCAGAGCTCAACACAGCAGAGCAATGACAAAAGTCAG	8940
QY	8949	CTGAGAACTGGTGGTATGAGCAAAAAAAGAAAGGCGACCTGATGTAAATCAATCCTAGA	9008
Db	8941	CTGAGAACTGGTGGTATGAGCAAAAAAAGAAAGGCGACCTGATGTAAATCAATCCTAGA	9000
QY	9009	TCCCTTACATGTGACTCCCTGCTCCGCAAAAGAGACTTGTCTACTGTGCTTCTTGTAG	9068
Db	9001	TCCCTTACATGTGACTCCCTGCTCCGCTCCGCAAAAGAGACTTGTCTACTGTGCTTCTTGTAG	9060
QY	9069	AGAAACACTAGAAACTGCCATGTCGTCAATAAGAGCTCTACCTGGAAACAAAAGTTGTT	9128
Db	9061	AGAAACACTAGAAACTGCCATGTCGTCAATAAGAGACTCTACCTGGAAACAAAAGTTGTT	9120
QY	9129	CTTTGGTGGTAACTACCTGGTGGGCGAGTGGTATACGCTCTTTAAATAATGACCAATACGT	9188
Db	9121	CTTTGGTGGTAACTACCTGGTGGGCGAGTGGTATACGCTCTTTAAATAATGACCAATACGT	9180
QY	9189	CTCACGCTATTTTAAATGATATGCTGTGATTTGTACACTTTTAAGTCCCTGACTT	9248
Db	9181	CTCACGCTATTTTAAATGATATGCTGTGATTTGTACACTTTTAAGTCCCTGACTT	9240
QY	9249	CATATTTGGCCTCATCTGATGTTCTTTGTGTTGGCCACACACACTGGTGAATAATGA	9308
Db	9241	CATATTTGGCCTCATCTGATGTTCTTTGTGTTGGCCACACACACTGGTGAATAATGA	9300
QY	9309	CATTTGCAGTATTTGTTTACTGACAGCGCTCTCGGGGCCATCCATGCAAGACACAT	9368
Db	9301	CATTTGCAGTATTTGTTTACTGACAGCGCTCTCGGGGCCATCCATGCAAGACACAT	9360
QY	9369	TAGTGTGGCAATGTTTTTTCACACTATATATATATGTCGTACTTTGTGTGAATAGAAATAC	9428
Db	9361	TAGTGTGGCAATGTTTTTTCACACTATATATATATATGTCGTACTTTGTGTGAATAGAAATAC	9420
QY	9429	TACAAAGATAGGCAAAAGTGTTCCTCAGTAGTAAAGCATATGGGTTTAACCTAAT	9488
Db	9421	TACAAAGATAGGCAAAAGTGTTCCTCAGTAGTAAAGCATATGGGTTTAACCTAAT	9480
QY	9489	TATCTCTCTATAAATATTTCTATATATATGATCATATAGTGACCTAATGTTATTC	9548
Db	9481	TATCTCTCTATAAATATTTCTATATATATGATCATATAGTGACCTAATGTTATTC	9540
QY	9549	TGCACCTTTTAAACCTTTTTTAAATTTACTCGTATATATGAAGCAGCCCTGATTTTAATTT	9608
Db	9541	TGCACCTTTTAAACCTTTTTTAAATTTACTCGTATATATGAAGCAGCCCTGATTTTAATTT	9600
QY	9609	AAAAAAAAAAAA 9618	
Db	9601	AAAAAAAAAAAA 9610	
RESULT 2			
AB011407			
LOCUS			
AB011407 431 bp mRNA linear BCT 27-JAN-1995			
DEFINITION			
Buchnera aphidicola mRNA for riboflavin synthase beta chain,			
partial cds.			
AB011407			
ACCESSION			
AB011407.1 GI:4190955			

KEYWORDS riboflavin synthase beta chain.
SOURCE Buchnera aphidicola (specific_host:Acyrthosiphon pisum) cDNA to mRNA.
ORGANISM Buchnera aphidicola
REFERENCE Bacteria: Proteobacteria; gamma subdivision; Buchnera.
AUTHORS 1 (bases 1 to 431)
TITLE Nakabachi, A.
JOURNAL Direct Submission
 Submitted (21-FEB-1998) Atsushi Nakabachi, University of Tokyo, Department of Biological Sciences, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan (E-mail: ss7722@hongo.ecc.u-tokyo.ac.jp, Tel:+81-3-3812-2111, Fax:+81-3-3816-1965)
REFERENCE 2 (sites)
AUTHORS Nakabachi, A. and Ishikawa, H.
TITLE Provision of riboflavin to the host aphid, Acyrthosiphon pisum, by endosymbiotic bacteria, Buchnera
JOURNAL J. Insect Physiol. 45, 1-6 (1999)
FEATURES Location/Qualifiers
 source 1..431
 /organism="Buchnera aphidicola"
 /specific_host="Acyrthosiphon pisum"
 /db_xref="taxon:9"
 <1..354
 /note="probable"
 /codon_start=1
 /transl_table=11
 /product="riboflavin synthase beta chain"
 /protein_id="BAA74435.1"
 /db_xref="GI:4190956"
 /translation="VHEHILKIVYGFVEIPTIASYIAKSGYDAIIAGTIIKQGT DHEFYKNDTSSLSRISNQYPLPTIGLTITKNIKESIERGRTKMGKNGSDAALAE MINVYMKIKKVIYY"
BASE COUNT 182 a 58 c 59 g 132 t
ORIGIN
 Query Match 0.2%; Score 24; DB 1; Length 431;
 Best Local Similarity 100.0%; Pred. No. 3.1;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3957 AAAACATGAGCAGTCATATTGAAA 3980
 |||||
Db 230 AAAACATGAGCAGTCATATTGAAA 253

RESULT 3
HUMTCBXN/C HUMTCBXN 495 bp DNA linear PRI 13-JAN-1995
LOCUS Human T-cell receptor aberrantly rearranged beta-chain V1-D2.1
DEFINITION gene.
ACCESSION M13575.1 GI:338940
VERSION D-region; T-cell receptor; T-cell receptor beta-chain; V-region;
KEYWORDS processed gene; pseudogene.
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 495)
AUTHORS Duby, A.D. and Seidman, J.G.
TITLE Abnormal recombination products result from aberrant DNA rearrangement of the human T-cell antigen receptor beta-chain gene
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 83 (13), 4890-4894 (1986)
MEDLINE 86259700
COMMENT The coding sequence of D2.1 has been deleted but its 5' flanking sequence was retained. This V segment appears to be a pseudogene because it has a nucleotide deletion at position 284 resulting in a frameshift mutation. Between V1 and D2.1 there are 19 nucleotides of unknown origin.
FEATURES Location/Qualifiers
 source 1..495
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_line="CEM"

gene /cell_type="T cell"
 join(15..63,152..>495)
 /gene="pseudo-TCB"
CDS join(15..63,152..>495)
 /gene="pseudo-TCB"
 /codon_start=1
 /pseudo
 64..151
 /number=1
exon 152..>495
 /gene="pseudo-TCB"
 /number=2
 441..442
 /gene="pseudo-TCB"
 /organism="Homo sapiens"
 /gene="pseudo-TCB"
 /organism="Homo sapiens"
misc_recomb /organism="Homo sapiens"
misc_recomb /organism="Homo sapiens"
BASE COUNT 103 a 135 c 132 g 124 t 1 others
ORIGIN
 Query Match 0.2%; Score 24; DB 9; Length 495;
 Best Local Similarity 100.0%; Pred. No. 3;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5855 GCAGAGAGAGCAAGCTGATGCT 5878
 |||||
Db 149 GCAGAGAGAGCAAGCTGATGCT 126

RESULT 4
HUMTCBZE/C HUMTCBZE 500 bp DNA linear PRI 13-JAN-1995
LOCUS Human T-cell receptor unproductively rearranged beta chain gene
DEFINITION V-region.
ACCESSION M31347
VERSION M31347.1 GI:339021
KEYWORDS T-cell receptor; V-region; processed gene; pseudogene.
SOURCE Human T-cell line HPB-ALL DNA, clone lambda HPB-2.
ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 500)
AUTHORS Ikuta, K., Ogura, T., Shimizu, A. and Honjo, T.
TITLE A joining-diversity-joining complex generated by inversion mechanism and a variable-diversity complex in the beta-chain gene of the human T-cell receptor
JOURNAL Nucleic Acids Res. 14 (12), 4899-4909 (1986)
MEDLINE 86259066
FEATURES Location/Qualifiers
 source 1..500
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="7q35"
 24..72
 /gene="TCRB"
 join(24..72,161..449)
 /partial
 /note="putative"
 24..72
 /gene="TCRB"
 /note="T-cell receptor beta chain V-region, exon 1;
 G00-120-405; putative"
 <24..72
 /gene="TCRB"
 /note="T-cell receptor beta chain V-region; G00-120-405"
 /pseudo
 /number=1
 161..>449
 /note="T-cell receptor beta chain V-region"
 /pseudo
 /number=2
 161..449
exon
exon


```

BASE COUNT      104 a      140 c      131 g      125 t
ORIGIN
Query Match      0.2%; Score 24; DB 9; Length 500;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5855 GCAGAGAGGACAAAGCTGATGCT 5878
Db 158 GCAGAGAGGACAAAGCTGATGCT 135

RESULT 5
LOCUS      HSDNATCRA/c      616 bp      DNA      linear      PRI 26-JAN-1995
DEFINITION H.sapiens TCR V Beta 13.2 gene (allele a).
ACCESSION X75418
VERSION X75418.1 GI:472946
KEYWORDS   TCR V-beta 13.2 gene.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 616)
AUTHORS   Kay, R.A., Snowden, N., Hajeer, A.H., Boylston, A.W. and Ollier, W.E.
TITLE     Importance of allelic variation outside the coding regions of the
          TCRBV13.2 gene
JOURNAL   Eur. J. Immunol. 24 (11), 2863-2867 (1994)
MEDLINE   95045939
REFERENCE  2 (bases 1 to 616)
AUTHORS   Kay, R.
TITLE     Direct Submission
JOURNAL   Submitted (07-OCT-1993) R. Kay, University of Manchester, ARC
          Epidemiology Research Unit, Stopford Bldg, Oxford Rd, Manchester
          M13 9PT, UK

FEATURES
source     1..616
            /organism="Homo sapiens"
            /isolate="patient 1"
            /db_xref="taxon:9606"
            /chromosome="7"
            /map="q32-q35"
            /cell_type="polymorph"
gene       1..616
            /gene="TCR V Beta 13.2 (allele a)"
            1..231
            /gene="TCR V Beta 13.2 (allele a)"
            232..281
            /gene="TCR V Beta 13.2 (allele a)"
            /number=1
            233..378
            /gene="TCR V Beta 13.2 (allele a)"
            /product="TCR V Beta 13.2"
            join(233..281,371..>616)
            /gene="TCR V Beta 13.2 (allele a)"
            /codon_start=1
            /product="TCR V Beta 13.2"
            /protein_id="CAA53170.1"
            /db_xref="GI:642237"
            /translation="MSIGLLCCGAFSLMAGPVNAGVYQTQPKFRVLTGQSMTLCAQ
            DMNHMYWYRQDPGKGLRLIHYSVGEITAKGEVDPDGYNVSRLKQNFLLGLE"
            282..370
            /gene="TCR V Beta 13.2 (allele a)"
            /number=1
            371..>616
            /gene="TCR V Beta 13.2 (allele a)"
            /number=2
            379..>616
            /gene="TCR V Beta 13.2 (allele a)"
            /product="TCR V Beta 13.2"

sig_peptide
CDS
intron
exon
mat_peptide

```

```

BASE COUNT      138 a      160 c      169 g      149 t
ORIGIN
Query Match      0.2%; Score 24; DB 9; Length 616;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5855 GCAGAGAGGACAAAGCTGATGCT 5878
Db 368 GCAGAGAGGACAAAGCTGATGCT 345

RESULT 6
LOCUS      HSTRVB132/c      737 bp      DNA      linear      PRI 05-APR-1992
DEFINITION H.sapiens TCR V-beta 13.2 gene for TCR beta chain variable region.
ACCESSION X61445
VERSION X61445.1 GI:37435
KEYWORDS   T-cell receptor; TCR beta chain variable region; TCR V-beta 13.2
          gene.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 737)
AUTHORS   Li, Y., Szabo, P. and Posnett, D.N.
TITLE     The genomic structure of human V beta 6 T cell antigen receptor
          genes
JOURNAL   J. Exp. Med. 174 (6), 1537-1547 (1991)
MEDLINE   92078864
REMARK     Erratum: [[published erratum appears in J Exp Med 1992 Feb
          1.175(2):617]]
          2 (bases 1 to 737)
REFERENCE  1.175(2):617]]
AUTHORS   Li, Y.
TITLE     Direct Submission
JOURNAL   Submitted (09-AUG-1991) Y. Li, Cornell Univ Medical College, 1300
          York Avenue, Room D-601, Box 56, New York 10021, NY, USA

FEATURES
source     1..737
            /organism="Homo sapiens"
            /isolate="human genomic library, Clontech labs., CAT#
            HU1067J"
            /db_xref="taxon:9606"
            /chromosome="7"
            /map="7q35"
            /clone="#5-2"
            /tissue_type="placental"
            /clone_lib="human genomic library"
            /germline
            1..266
            267..315
            /gene="TCR V-beta 13.2"
            /note="first leader"
            /number=1
            join(267..315,404..698)
            /gene="TCR V-beta 13.2"
            join(267..315,404..698)
            /partial
            /gene="TCR V-beta 13.2"
            /codon_start=1
            /protein_id="CAA43685.1"
            /db_xref="GI:37436"
            /translation="MSIGLLCCGAFSLMAGPVNAGVYQTQPKFRVLTGQSMTLCAQ
            DMNHMYWYRQDPGKGLRLIHYSVGEITAKGEVDPDGYNVSRLKQNFLLGLESAAP
            SQTSVYFCASSTYS"
            316..403
            /gene="TCR V-beta 13.2"
            /number=1
            404..698
            /gene="TCR V-beta 13.2"
            /note="second leader and V region"
            /number=2
            170 a      195 c      189 g      183 t

```

ORIGIN	
Query Match	0.2%; Score 24; DB 9; Length 737;
Best Local Similarity	100.0%; Pred. No. 2.9;
Matches	24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
0Y	5855 GCAGAGAAGCAAAAGCTGATGCT 5878
Db	401 GCAGAGAAGCAAAAGCTGATGCT 378
RESULT 7	
LOCUS	HSU07976 8045 bp DNA linear PRI 16-OCT-1994
DEFINITION	Human T cell receptor beta (TCRBV7S2, TCRBV13S2-1, TCRBV6S7-1)
ACCESSION	U07976
VERSION	U07976.1 GI:558422
KEYWORDS	
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 8045) Zhao,T.M., Whitaker,S.E. and Robinson,M.A. A genetically determined insertion/deletion related polymorphism in human T cell receptor beta chain (TCRB) includes functional variable gene segments J. Exp. Med. 180, 1405-1414 (1994)
AUTHORS	95016432 2 (bases 1 to 8045) Robinson,M.A. Direct Submission Submitted (24-MAR-1994) Mary Ann Robinson Ph.D., Laboratory of Immunogenetics, NIDDK/National Institutes of Health, 12441 Parklawn Dr., Room 79, Rockville, MD 20852, USA
JOURNAL	Location/Qualifiers
MEDLINE	1. 8045
REFERENCE	/organism="Homo sapiens"
AUTHORS	/db_xref="taxon:9606"
TITLE	/chromosome="7"
JOURNAL	/map="7q35"
	/clone="D/S"
	/haplotype="TCRBV deleted 2"
	/sex="female"
	/cell_line="SCGMV"
	/cell_type="B cell"
	join(311. .359,469. .>758)
gene	/gene="TCRBV7S2"
	join(311. .359,469. .>758)
CDS	/gene="TCRBV7S2"
	/codon_start=1
	/product="T cell receptor beta"
	/protein_id="AAA50585.1"
	/db_xref="GI:558423"
	/translation="MGCRLCCAVLCGAVPMETGVOTPRHLVMGMTNKSLSKCEQ HIGNAYMYKQSAKPKLEIMFYVNFKEQENNSVPSRFSPENSSHLFLHTLTQDP EDSLVLYKASS"
	360. 468
intron	/gene="TCRBV7S2"
	join(4000. .4048,4137. .>4426)
gene	/gene="TCRBV13S2-1"
	join(4000. .4048,4137. .>4426)
CDS	/gene="TCRBV13S2-1"
	/codon_start=1
	/product="T cell receptor beta"
	/protein_id="AAA50586.1"
	/db_xref="GI:558424"
	/translation="MSDGLCCGAFSLTMAGPVNAGVOTPKFVILKTGQSMFLCAQ DMNEEYMYWRODPGGMGLRLIHYSVGGTTAKGVPDGYNVSRLKKNPFLGLGSAAAP SQRSVLYKASS"
	4049. .4136
intron	/gene="TCRBV13S2-1"

```

gene
  join(7627..7675,7824..>8045)
  /gene="TCRBV657-1"
  join(7627..7675,7824..>8045)
  /gene="TCRBV657-1"
  /codon_start=1
  /product="T cell receptor beta"
  /protein_id="AAAS0587.1"
  /db_xref="GI:558425"
  /translation="MGTRLFVWAFCLIGADHTGAGVSQSPSNKVTXKQDELRCDF
ISGHTALWYROSIGQGLEFLFYFGNSAPDKSLPSPRFAERTG"
  7676..7823
  /gene="TCRBV657-1"

intron
  /gene="TCRBV657-1"

BASE COUNT      2256 a      1813 c      1809 g      2167 t
ORIGIN
Query Match      0.2%; Score 24; DB 9; Length 8045;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5855 GCAGAGAGAGCAACAAAGCTGATGCT 5878
      |||||||
Db 4134 GCAGAGAGAGCAACAAAGCTGATGCT 4111

RESULT 8
HSU07978/c 33037 bp DNA linear PRI 16-OCT-1994
LOCUS Human T cell receptor beta (TCRBV7S2, TCRBV13S2-1b, TCRBV9S2p,
DEFINITION TCRBV7S3, TCRBV13S2-1b, TCRBV6S7-1) genes, TCRBV inserted
haplotype, partial cds.
VERSION U07978
KEYWORDS U07978.1 GI:558429
SOURCE
  human.
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
  1 (bases 1 to 33037)
  Zhao,T.M., Whitaker,S.E. and Robinson,M.A.
  A genetically determined insertion/deletion related polymorphism in
  human T cell receptor beta chain (TCRB) includes functional
  variable gene segments
  J. Exp. Med. 180, 1405-1414 (1994)
  95016432
  2 (bases 1 to 33037)
  Robinson,M.A.
  Direct Submission
  Submitted (24-MAR-1994) Mary Ann Robinson Ph.D., Laboratory of
  Immunogenetics, NIAID/National Institutes of Health, 12441 Parklawn
  Dr., Room 79, Rockville, MD 20852, USA
  Location/Qualifiers
    1..33037
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="7"
    /map="7q35"
    /clone="17s,17l,19,19p2,19q3,19q4"
    /haplotype="TCRBV inserted"
    /sex="male"
    /cell_line="SBOR"
    /cell_type="B cell"
    /join(3734..3782,3892..>4181)
    /gene="TCRBV7S2"
    /join(3734..3782,3892..>4181)
    /gene="TCRBV7S2"
    /codon_start=1
    /product="T cell receptor beta"
    /protein_id="AAAS0590.1"
    /db_xref="GI:558430"
    /translation="MGCRLLCCAVILGAVPMETGYTPRRHLVGMGTNKKSLKQSO
HLGNMAYWYQSAKKPLELFLFYVNFKEQLENNVSPSRFSPECNNSHLPLHLHTLDP
EDSLALYICASS"
    3783..3891
intron

```

```

gene      /gene="TCRBV7S2"
          join(7424..7472,7561..>7850)
          /gene="TCRBV13S2-1a"
          join(7424..7472,7561..>7850)
CDS       /gene="TCRBV13S2-1a"
          /codon_start=1
          /product="T cell receptor beta"
          /protein_id="AA50591.1"
          /db_xref="GI:558431"
          /translation="MSGLCCGAFSLMAGPVNAGVTQTPEKFRVLKGTQSMILLCAQ
DNHHEMTWYRDPDGMGLRLIHYSVGEETTAKEVDPDGYNVSLKKONFLGLESAAP
SOTSVYFCASS"
          7473..7560
CDS       /gene="TCRBV13S2-1a"
          join(17328..17376,17494..>17783)
          /note="TCRBV9S2p"
          /codon_start=1
          /pseudo
          17377..17493
          /note="TCRBV9S2p"
          join(20774..20822,20932..>21221)
          /gene="TCRBV7S3"
          join(20774..20822,20932..>21221)
CDS       /gene="TCRBV7S3"
          /codon_start=1
          /product="T cell receptor beta"
          /protein_id="AA50592.1"
          /db_xref="GI:558432"
          /translation="MGCRLCCAVILCGAVPMETGYTOTPRHLVGMTKKSICEQ
HGHNAWYKOSAKKPLELMFYSLDERVNNVPSRFSPECNSSHLPLHLTLTP
EDSALYLCASS"
          20823..20931
          /gene="TCRBV7S3"
          join(28984..29032,29121..>29410)
          /gene="TCRBV13S2-1b"
          join(28984..29032,29121..>29410)
CDS       /gene="TCRBV13S2-1b"
          /codon_start=1
          /product="T cell receptor beta"
          /protein_id="AA50593.1"
          /db_xref="GI:558433"
          /translation="MSLGLCCGAFSLMAGPVNAGVTOTPRKFRVLKGTQSMILLCAQ
DNHHEMTWYRDPDGMGLRLIHYSVGEETTAKEVDPDGYNVSLKKONFLGLESAAP
SOTSVYFCASS"
          29033..29120
          /gene="TCRBV13S2-1b"
          join(32617..32665,32816..>33037)
          /gene="TCRBV6S7-1"
          join(32617..32665,32816..>33037)
CDS       /gene="TCRBV6S7-1"
          /codon_start=1
          /product="T cell receptor beta"
          /protein_id="AA50594.1"
          /db_xref="GI:558434"
          /translation="MGRILFWAVCLIGADHTGAGVSQSPSKYTKRKDEVLKCDP
ISGHTALWYRSLGGLEFLIFQGNAPSKGLSDFNSAERTG"
          32666..32815
          /gene="TCRBV6S7-1"
          join(7118..7074..9476..t
BASE COUNT 9369 a 7118 c 7074 g 9476 t
ORIGIN
Query Match      0.28; Score 24; DB 9; Length 33037;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5855 GCAGAGAAGCAAAAGCTGATGCT 5878
      |||||||
Db 29118 GCAGAGAAGCAAAAGCTGATGCT 29095
      |||||||
RESULT 9
AC101336
LOCUS      AC101336      72007 bp      DNA      linear      HTG 23-NOV-2001

```

```

DEFINITION Mus musculus clone RP23-109p22, LOW-PASS SEQUENCE SAMPLING.
AC101336
AC101336.1 GI:17060111
HTG: HTGS PHASE0.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 72007)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
JOURNAL
REFERENCE
2 (bases 1 to 72007)
Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
Anderson, S., Barta, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fairo, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., Larocque, K.,
Lamaze, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Melidim, J.,
Meneus, L., Minova, T., Menga, V., Murphy, T., O'Donnell, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Neill, J., O'Neill, D.,
Oliver, J., Peterson, R., Phunkhaph, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainou, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 109_P_22
Center clone name: 109_P_22
-----
* NOTE: This record contains 89 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1
* 715 814: gap of 100 bp
* 815 1492: contig of 678 bp in length
* 1493 1592: gap of 100 bp
* 1593 2314: contig of 722 bp in length
* 2315 2414: gap of 100 bp
* 2415 3136: contig of 722 bp in length
* 3137 3236: gap of 100 bp
* 3237 3958: contig of 722 bp in length
* 3959 4058: gap of 100 bp
* 4059 4794: contig of 736 bp in length
* 4795 4894: gap of 100 bp

```

* 4895 5605: contig of 711 bp in length
* 5606 5705: gap of 100 bp
* 5706 6407: contig of 702 bp in length
* 6408 6507: gap of 100 bp
* 6508 7233: contig of 726 bp in length
* 7234 7333: gap of 100 bp
* 7334 8036: contig of 703 bp in length
* 8037 8136: gap of 100 bp
* 8137 8853: contig of 717 bp in length
* 8854 8953: gap of 100 bp
* 8954 9669: contig of 716 bp in length
* 9670 9769: gap of 100 bp
* 9770 10476: contig of 707 bp in length
* 10477 10576: gap of 100 bp
* 10577 11272: contig of 696 bp in length
* 11273 11372: gap of 100 bp
* 11373 12071: contig of 699 bp in length
* 12072 12171: gap of 100 bp
* 12172 12861: contig of 680 bp in length
* 12862 12961: gap of 100 bp
* 12962 13685: contig of 724 bp in length
* 13686 13785: gap of 100 bp
* 13786 14509: contig of 724 bp in length
* 14510 14609: gap of 100 bp
* 14610 15322: contig of 713 bp in length
* 15323 15422: gap of 100 bp
* 15423 16132: contig of 710 bp in length
* 16133 16232: gap of 100 bp
* 16233 16949: contig of 717 bp in length
* 16950 17049: gap of 100 bp
* 17050 17745: contig of 686 bp in length
* 17746 17845: gap of 100 bp
* 17846 18505: contig of 660 bp in length
* 18506 18605: gap of 100 bp
* 18606 19309: contig of 704 bp in length
* 19310 19409: gap of 100 bp
* 19410 20122: contig of 713 bp in length
* 20123 20222: gap of 100 bp
* 20223 20928: contig of 706 bp in length
* 20929 21028: gap of 100 bp
* 21029 21756: contig of 728 bp in length
* 21757 21856: gap of 100 bp
* 21857 22581: contig of 725 bp in length
* 22582 22681: gap of 100 bp
* 22682 23407: contig of 726 bp in length
* 23408 23507: gap of 100 bp
* 23508 24215: contig of 708 bp in length
* 24216 24315: gap of 100 bp
* 24316 25032: contig of 717 bp in length
* 25033 25132: gap of 100 bp
* 25133 25839: contig of 707 bp in length
* 25840 25939: gap of 100 bp
* 25940 26622: contig of 683 bp in length
* 26623 26722: gap of 100 bp
* 26723 27423: contig of 701 bp in length
* 27424 27523: gap of 100 bp
* 27524 28229: contig of 706 bp in length
* 28230 28329: gap of 100 bp
* 28330 29058: contig of 729 bp in length
* 29059 29158: gap of 100 bp
* 29159 29866: contig of 708 bp in length
* 29867 29966: gap of 100 bp
* 29967 30671: contig of 705 bp in length
* 30672 30771: gap of 100 bp
* 30772 31511: contig of 740 bp in length
* 31512 31611: gap of 100 bp
* 31612 32331: contig of 720 bp in length
* 32332 32431: gap of 100 bp
* 32432 33140: contig of 709 bp in length
* 33141 33240: gap of 100 bp
* 33241 33957: contig of 717 bp in length
* 33958 34057: gap of 100 bp
* 34058 34771: contig of 714 bp in length

* 34772 34871: gap of 100 bp
* 34872 35577: contig of 706 bp in length
* 35578 35677: gap of 100 bp
* 35678 36370: contig of 693 bp in length
* 36371 36470: gap of 100 bp
* 36471 37165: contig of 695 bp in length
* 37166 37265: gap of 100 bp
* 37266 37980: contig of 715 bp in length
* 37981 38080: gap of 100 bp
* 38081 38793: contig of 713 bp in length
* 38794 38893: gap of 100 bp
* 38894 39611: contig of 718 bp in length
* 39612 39711: gap of 100 bp
* 39712 40417: contig of 706 bp in length
* 40418 40517: gap of 100 bp
* 40518 41227: contig of 710 bp in length
* 41228 41327: gap of 100 bp
* 41328 42047: contig of 720 bp in length
* 42048 42147: gap of 100 bp
* 42148 42820: contig of 673 bp in length
* 42821 42920: gap of 100 bp
* 42921 43614: contig of 694 bp in length
* 43615 43714: gap of 100 bp
* 43715 44439: contig of 725 bp in length
* 44440 44539: gap of 100 bp
* 44540 45261: contig of 722 bp in length
* 45262 45361: gap of 100 bp
* 45362 46094: contig of 733 bp in length
* 46095 46194: gap of 100 bp
* 46195 46884: contig of 660 bp in length
* 46885 46984: gap of 100 bp
* 46985 47688: contig of 704 bp in length
* 47689 47788: gap of 100 bp
* 47789 48511: contig of 723 bp in length
* 48512 48611: gap of 100 bp
* 48612 49329: contig of 718 bp in length
* 49330 49429: gap of 100 bp
* 49430 50167: contig of 738 bp in length
* 50168 50267: gap of 100 bp
* 50268 50980: contig of 713 bp in length
* 50981 51080: gap of 100 bp
* 51081 51787: contig of 707 bp in length
* 51788 51887: gap of 100 bp
* 51888 52585: contig of 688 bp in length
* 52586 52685: gap of 100 bp
* 52686 53393: contig of 708 bp in length
* 53394 53493: gap of 100 bp
* 53494 54209: contig of 716 bp in length
* 54210 54309: gap of 100 bp
* 54310 55028: contig of 719 bp in length
* 55029 55128: gap of 100 bp
* 55129 55843: contig of 715 bp in length
* 55844 55943: gap of 100 bp

Query Match 0.2%: Score 24; DB 2; Length 72007;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5856 CAGAGAGACAAGCTGATGCTG 5879
|||||
Db 34396 CAGAGAGACAAGCTGATGCTG 34419

RESULT 10
AC022170
LOCUS
DEFINITION Homo sapiens chromosome 2 clone RP11-2001, WORKING DRAFT SEQUENCE,
3 unordered pieces.
ACCESSION AC022170
VERSION AC022170.14 GI:13122768
KEYWORDS
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULTOP; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 155021)
Abola, A.P., Bruno, D., Conn, L., Dela Rosa, M., Faulkner, D., Federspiel, N., Glukhov, S., Hansen, N., Herman, Z.S., Hyman, R., Mao, J., Lam, B., Marathe, R., Miranda, M., Morehouse, A.J., Nguyen, M., Oefner, P., Palm, C.J., Ramirez, D., Southwick, A.M., Wilhelmy, J., Yu, S., and Davis, R.W.
Unpublished
2 (bases 1 to 155021)
Bruno, D., Conn, L., Dela Rosa, M., Federspiel, N., Foreman, P., Glukhov, S., Hansen, N., Hyman, R., Mao, J., Marathe, R., Morehouse, A.J., Oefner, P., Palm, C.J., Ramirez, D., Wilhelmy, J., Yu, S., and Davis, R.W.
Direct Submission
Submitted (26-JAN-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
On Feb 24, 2001 this sequence version replaced gi:12863205.
----- Genome Center
Center: Stanford DNA Sequencing and Technology Development Center
Center code: SDSLDC
Web site: <http://sequence-www.stanford.edu/group/human/>
Contact: hum-info@sequence.stanford.edu
----- Project Information
Center project name: 768
Center clone name: RP11-2061
----- Summary Statistics
Sequencing Vector: M13mp18; X02513; 98% of reads
Sequencing Vector: plasmid; plasmid_accession: 2% of reads
Chemistry: Dye-terminator Big Dye; 92% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 153360 bases at least Q40
Consensus quality: 154052 bases at least Q20
Insert size: 11631; agarose-fp
Insert size: 154821; sum-of-contigs
Quality coverage: 12.3x in Q20 bases; agarose-fp
Quality coverage: 9.3x in Q20 bases; sum-of-contigs
* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
* 1 20875: contig of 20875 bp in length
* 20876 20975: gap of unknown length
* 20976 52837: contig of 31862 bp in length
* 52838 52937: gap of unknown length
* 52938 155021: contig of 102084 bp in length.
Location/Qualifiers
1. 155021
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-2061"
/clone_11p="RPC1 human BAC library 11"
1. 20875
/note="assembly_name:Contig33
clone_end:77"
misc_feature 20976..52837
/note="assembly_name:Contig34
clone_end:SP6"
52938..155021
/note="assembly_name:Contig35"
BASE COUNT 45493 a 33367 c 32844 g 43116 t 201 others
ORIGIN
Query Match 0.2% Score 24; DB 2; Length 155021;

Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3390 ACACAAACATGACTTGACGAA 3413
DB 142072 ACACAAACATGACTTGACGAA 142095
RESULT 11
AC073415/c 215977 bp DNA linear HTG 24-JAN-2002
LOCUS Homo sapiens chromosome 2 clone RP11-84C2, WORKING DRAFT SEQUENCE,
DEFINITION 2 unordered pieces.
ACCESSION AC073415.5 GI:18308794
VERSION AC073415.5
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 215977)
The sequence of Homo sapiens clone
2 (bases 1 to 215977)
Waterston, R.H.
Direct Submission
Submitted (16-JUN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
On Jan 24, 2002 this sequence version replaced gi:16195290.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: H.NH0084C02
----- Summary Statistics -----
Sequencing vector: M13; 42%
Sequencing vector: plasmid; 58%
Chemistry: Dye-terminator Big Dye; 42% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 214602 bases at least Q40
Consensus quality: 215177 bases at least Q30
Consensus quality: 215466 bases at least Q20
Insert size: 227000; agarose-fp
Insert size: 217039; sum-of-contigs
Quality coverage: 9.32 in Q20 bases; agarose-fp
Quality coverage: 8.26 in Q20 bases; sum-of-contigs
* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
* 1 28725: contig of 28725 bp in length
* 28726 28825: gap of unknown length
* 28826 215977: contig of 187152 bp in length.
Location/Qualifiers
1. 215977
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-84C2"
1. 28725
/note="assembly_name:Contig50
misc_feature 215977
FEATURES
source
misc_feature 215977
/note="assembly_name:Contig50

misc.feature clone_end:SP6
 vector_side:right"
28826. .215977
/note="assembly_name:Contig51
clone_end:T7
vector_side:right"
BASE COUNT 61078 a 40739 c 43688 g 70372 t 100 others
ORIGIN

Query Match 0.24; Score 24; DB 2; Length 215977;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3390 ACACAAACATGACTGACAGAA 3413
Db 161277 ACACAAACATGACTGACAGAA 161254
|||||

RESULT 12
AX336388/c 267156 bp DNA linear PAT 09-JAN-2002
LOCUS
DEFINITION Sequence 6897 from Patent WO0194629.
ACCESSION AX336388
VERSION AX336388.1 GI:18127107
KEYWORDS

SOURCE
ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
REFERENCE
AUTHORS Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
 Horrigan,S., Soppet,D.R. and Weaver,Z.
 Cancer gene determination and therapeutic screening using signature
 gene sets
 Patent: WO 0194629-A 6897 13-DEC-2001;
 Avalon Pharmaceuticals (US)
 location/Qualifiers
 1.267156
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
BASE COUNT 76527 a 56343 c 55787 g 78499 t
ORIGIN

Query Match 0.24; Score 24; DB 6; Length 267156;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5855 GCAGAGAAGCAGCAAGCTGCT 5878
Db 163714 GCAGAGAAGCAGCAAGCTGCT 163691
|||||

RESULT 13
U66059/c 267156 bp DNA linear PRI 30-NOV-1998
LOCUS
DEFINITION Human germline T-cell receptor beta chain
 Dopamine-beta-hydroxylase-like, TRY1, TRY2, TRY3, TCRBV27S1P,
 TCRBV22S1A2N1T, TCRBV9S1A1T, TCRBV51A1N2T, TCRBV51A1T, TCRBV13S3,
 TCRBV6S7P, TCRBV7S3A2T, TCRBV13S2A1T, TCRBV9S2A2PT, TCRBV7S2A1N4T,
 TCRBV13S9/13S2A1T, TCRBV6S5A1N1, TCRBV50S1P, TCRBV13S1, TCRBV13S5,
 TCRBV6S1A1N1, TCRBV13S2S1P, TCRBV5S5P, TCRBV1S1A1N1, TCRBV12S2A1T,
 TCRBV21S1, TCRBV8S4P, TCRBV12S3, TCRBV21S3A2N2T, TCRBV8S5P,
 TCRBV13S1 genes from bases 1 to 267156 (section 1 of 3).
U66059.1 GI:1552494
C region: C-beta gene segment; D region: J-beta gene segment;
J-segment: T-cell receptor beta chain; TCR-beta gene; V-beta gene
segment; V-segment: cell membrane protein; constant region;
diversity region: germline; joining segment; trypsin; trypsinogen;
variable segment.
Homo sapiens (clone: K41A) (clone.lib: CGM1:YAC D49H4) DNA: Homo
sapiens (clone: K35) (clone.lib: CGM1:YAC D49H4) DNA: Homo sapiens (clone:
(clone: K26) (clone.lib: CGM1:YAC D49H4) DNA: Homo sapiens (clone:

K56) (clone.lib: CGM1:YAC D49H4) DNA: Homo sapiens (library:
ATCC1475) (clone: K21B) (clone.lib: Kai Wang's) DNA: Homo sapiens
(library: ATCC1475) (clone: G54) (clone.lib: Kai Wang's) DNA: Homo
sapiens (library: Helai) (clone: H137) (clone.lib: Eric Lai's) DNA:
Homo sapiens (library: sperm) (clone: H18) (clone.lib: Eric Lai's)
DNA: Homo sapiens (clone: H18/G15gap) DNA: Homo sapiens (library:
ATCC1475) (clone: G15) (clone.lib: Kai Wang's) DNA: Homo sapiens
(library: ATCC1475) (clone: X1A) (clone.lib: Kai Wang's) DNA: Homo
sapiens (clone: A27) (clone.lib: CGM1:YAC 234 A6F6) DNA: Homo
sapiens (clone: A212partial) (clone.lib: CGM1:YAC 234 A6F6) DNA:
Homo sapiens (clone: A14) (clone.lib: CGM1:YAC 234 A6F6) DNA: Homo
sapiens (library: sperm) (clone: H7.1) (clone.lib: Eric Lai's) DNA:
Homo sapiens (clone: H12.18) (clone.lib: Eric Lai's) DNA: Homo
sapiens (library: Helai) (clone: H130) (clone.lib: Eric Lai's) DNA:
Homo sapiens (clone: A16) (clone.lib: CGM1:YAC 234 A6F6) DNA: Homo
sapiens (clone: C215) (clone.lib: CGM1:YAC 234 A72B3) DNA: Homo
sapiens (clone: G1) (library: ATCC1475) (clone.lib: Kai Wang's)
DNA: Homo sapiens (clone: C68) (clone.lib: CGM1:YAC 210 A38G1)
DNA: Homo sapiens (clone: C21) (clone.lib: CGM1:YAC 234 A72B3)
DNA: Homo sapiens (library: ATCC1475) (clone: X11) (clone.lib: Kai
Wang's) DNA: Homo sapiens (library: ATCC1475) (clone: X6A)
(clone.lib: Kai Wang's) DNA: Homo sapiens (library: ATCC1475)
(clone: CBG1) (clone.lib: Kai Wang's) DNA: Homo sapiens (clone:
CBG1/C29gap) DNA: and Homo sapiens (library: ATCC1475) (clone: C29)
(clone.lib: Kai Wang's) DNA.

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
1 (bases 1 to 267156)
Rowen,L., Koop,B.F. and Hood,L.
The complete 685-kilobase DNA sequence of the human beta T cell
receptor locus
Science 272 (3269), 1755-1762 (1996)
96256474
2 (bases 1 to 267156)
Rowen,L., Seto,J., Smit,A., Acharya,C., Ahearn,M.E., Ankener,M.,
Baskin,D., Bumgarner,R., Chen,L., Chen,N., Deshpande,P., Faust,J.,
Howard,S., Jerome,N., Koop,B.F., Lee,H., Loretz,C., Paeger,B.,
Zackroff,K. and Hood,L.
Sequence determination of the human T cell receptor beta locus:
Strategy and error analysis
Unpublished
3 (bases 1 to 267156)
Rowen,L., Wang,K., Boyse,C., Ahearn,M.E., Charmley,P., Paeger,B.,
Lee,I., Chen,L., Trask,B., Nickerson,D., Seto,D. and Hood,L.
Sequence variation among several haplotypes in the human T cell
receptor beta locus
Unpublished
4 (bases 1 to 267156)
Rowen,L.
Direct Submission
Submitted (15-OCT-1994) L.Rowen leetowen@u.washington.edu
5 (bases 1 to 267156)
Rowen,L.
Direct Submission
Submitted (24-JUN-1997)
This sequence overlaps section 2 of the T cell receptor beta locus,
Genbank Accession Number U66060, by 9647 bases.
location/Qualifiers
1.267156
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="7q35"
complement(1..240)
/note="putative"
/rpt_family="Alu"
1..37031
/note="(vector PWE15A)"
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="K41A"
/cell_line="CGM1, haplotype B"
/clone.lib="YAC D49H4"

```

repeat_region complement(577, 709)
                /germline
                /note="SINE; putative"
                /rpt_family="MIR"
repeat_region complement(1463, 1575)
                /note="MAlR retroposon LTR; putative"
                /rpt_family="MTIG"
                2307, 2428
repeat_region /note="DNA transposon fossil; putative"
                /rpt_family="MER3; fragment 1"
                2429, 7244
repeat_region /note="LINE; putative"
                /rpt_family="L1PA7; fragment 1"
                2429, 2988
repeat_region /note="putative"
                /rpt_family="LINE 1"
                7245, 7975
repeat_region /note="LINE; putative"
                /rpt_family="L1PA7; fragment 2"
                7997, 8294
repeat_region /note="putative"
                /rpt_family="Alu"
                8295, 8463
repeat_region /note="putative"
                /rpt_family="MER3"
                8764, 8793
repeat_region /note="15 bp tandem duplication; putative"
                /rpt_type=tandem
                complement(9085, 9224)
repeat_region /note="DNA transposon fossil; putative"
                /rpt_family="MER3"
                9888, 9965
repeat_region /note="SINE; putative"
                /rpt_family="MIR"
                10076, 10320
repeat_region /note="LINE; putative"
                /rpt_family="L1MA4; fragment 1"
                10321, 10616
repeat_region /note="putative"
                /rpt_family="Alu"
                10617, 10747
repeat_region /note="LINE; putative"
                /rpt_family="L1MA4; fragment 2"
                12541, 13965
repeat_region /note="60 (120) mer, 19x repeated, 91% conserved first
                5.5 120 mers; putative"
                /rpt_type=tandem
                15873, 16381
repeat_region /note="DNA transposon fossil; putative"
                /rpt_family="MER64"
                complement(19303, 19385)
                /note="SINE; putative"
                /rpt_family="MIR"
                19542, 19632
repeat_region /note="SINE; putative"
                /rpt_family="MIR"
                19682, 19755
repeat_region /note="LINE; putative"
                /rpt_family="L1MA3"
                19945, 20437
repeat_region /note="retroposon LTR; putative"
                /rpt_family="MER65B"
                complement(21066, 21541)
repeat_region /note="retroposon LTR; putative"
                /rpt_family="MER60A"
                complement(22822, 23202)
repeat_region /note="LINE; putative"
                /rpt_family="L1PA16"
                23444, 23601
repeat_region /note="SINE; putative"
                /rpt_family="MIR"
                complement(23615, 23909)
repeat_region /note="putative"

```

```

repeat_region /rpt_family="Alu"
                24043, 24079
                /note="SINE; putative"
                /rpt_family="MIR"
repeat_region /note="endogenous retroviral LTR; putative"
                /rpt_family="LTR1; fragment 1"
                complement(25086, 25502)
                /note="endogenous retroviral LTR; putative"
                /rpt_family="LTR1; fragment 2"
                complement(25975, 26233)
                /note="putative"
                /rpt_family="Alu"
                27165, 28630
repeat_region /note="LINE; putative"
                /rpt_family="L1P5; fragment 1"
                28056, 28507
                /note="putative"
                /rpt_family="LINE 1"
                28631, 28702
repeat_region /note="LINE; putative"
                /rpt_family="L1P5; fragment 2"
                complement(28713, 29101)
                /note="LINE; putative"
                /rpt_family="MIR2"
                complement(30772, 31746)
                /note="LINE; putative"
                /rpt_family="MER42C"
                complement(31996, 32116)
                /note="DNA transposon fossil; putative"
                /rpt_family="MER5B"
                32313, 32419
repeat_region /note="LINE; putative"
                /rpt_family="MER42C"
                complement(33276, 33406)
                exon

```

```

Query Match      0.2%  Score 24; DB 9; Length 267156;
Best Local Similarity 100.0%; Pred. NO. 1.8;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5855 GCAGAGAGGACAAAGCTGATGCT 5878
    |||
DB 163714 GCAGAGAGGACAAAGCTGATGCT 163691

```

```

RESULT 14
AP001119      293181 bp   DNA   linear   BCT 07-SBP-2000
DEFINITION   Buchnera sp. APS genomic DNA, complete sequence, segment 2/2.
ACCESSION    AP001119 AP000398 BA000003
VERSION      AP001119.1 GI:10038987
KEYWORDS     SOURCE
              Buchnera sp. APS (strain:APS isolate:Tokyo1998,
              specific_host:Acyrthosiphon pisum) DNA.
ORGANISM     Buchnera sp. APS
              Bacteria; Proteobacteria; gamma subdivision; Buchnera.
REFERENCE    1 (sites)
AUTHORS      Shigenobu,S., Watanabe,H., Hattori,M., Sakaki,Y. and Ishikawa,H.
TITLE        Genome sequence of the endocellular bacterial symbiont of aphids
              Buchnera sp. APS
JOURNAL      Nature 407 (6800), 81-86 (2000)
MEDLINE      20445173
REFERENCE    2 (bases 1 to 293181)
AUTHORS      Shigenobu,S., Hattori,M., Sakaki,Y. and Ishikawa,H.
TITLE        Direct Submission
JOURNAL      Submitted (24-JAN-2000) Shuji Shigenobu, University of Tokyo,
              Department of Biological Sciences, 7-3-1 Honjo, Bunkyo-ku, Tokyo
              113-0033, Japan (E-mail:shige@sc.riken.go.jp,
              URL:http://buchnera.gsc.riken.go.jp, Tel:81-3-5800-3553,
              Fax:81-3-5800-3553)
COMMENT      Collaboration Information:
              RIKEN, Genomic Sciences Center (GSC)

```

Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan
Tel:81-42-778-9923
FAX:81-42-778-9924
AP000398: Submitted(23-Aug-1999).

FEATURES
source Location/Qualifiers

1..293181
/organism="Buchnera sp. APS"
/strain="APS"
/isolate="Tokyo1998"
/specific_host="Acyrthosiphon pisum"
/db_xref="taxon:107806"
/note="primary endosymbiont of Acyrthosiphon pisum"
321..1280
/gene="crxb"
321..1280
/gene="crxb"
/EC_number="1.6.4.5"
/note="BU314"
/codon_start=1
/transl_table=11
/product="thioredoxin reductase"
/protein_id="BAB13022.1"
/db_xref="GI:10038988"
/translation="MDKVRHKKIILIGSPACGTAAIYARANDLPFLITGKNGCOL
MNTNIEWMPGDYKISGSELMNRMKHAIEKTYVICDPTVISVNFKNPFLLIGBN
KYTADSVIATGPNRYLGLQSELSFKGSGVTCACDGFYKNKEVAAGGNTAIE
ETLYSNFVKKYLHRCINFRAEKILDLREKIKSOKIITYLSINVAIIGNSGV
TALLIEOKNSKEKTESKIOVSGLFVAIGYTPNTNIFNKLKMDGIIQYTRQEHGNYT
OTSTGIFACGVIDHYVRYQAITSSASCGMALDSERYINSLV"
1387..1605
/gene="infa"
1387..1605
/gene="infa"
/note="BU315"
/codon_start=1
/transl_table=11
/product="translation initiation factor IF-1"
/protein_id="BAB13023.1"
/db_xref="GI:10038989"
/translation="MSKEENIMOGVVIDTLPNTWFRVLEKNHITYAHISGMKRNKY
IRLLGDKVTVELTPYDLTKGRIIFRSR"
1774..3534
/gene="asps"
1774..3534
/gene="asps"
/EC_number="6.1.1.12"
/note="BU316"
/codon_start=1
/transl_table=11
/product="aspartyl-tRNA synthetase"
/protein_id="BAB13024.1"
/db_xref="GI:10038990"
/translation="MRTKCYGNIIRISHNNKKKYLDCGWVHKVNLGCFITVDKRDYTCGL
VOVIELKNYKIEFKALNDRNEFCIQVFQVOKREKKNNIKIRIGEIEILANVLNLT
NTSKPLNFGENDSDSRKLYDLDSFLELNKIKNTYILIRNMTKNEFDI
ETPLTKSTPEGARDYLVPSRNHYGKFALPQSPOLKQILMISGIDRRYQVKKFERD
EDLRSDRPEFTQIDIEVSEFSAKKIRNLVENLIRKIMEIRINILNKKRPOISPHFAM
KRGSDKRDLRNPIITIDVSNTEFKDKKTSFENLNLKKNRRIALICISGAHSRKKI
DDYTRYVRFDAKKLFYMKIKCKLGCGLHSISTKNILDEIILKEIEKSQSKNDIL
FLIADQEHIVNKSGLMRKIGIDILNITKKNWEPDLWVNFPEKDDIOGNISVHPH
FTVAQNMREILKNSPDLAISDYDLIINGEYIGGVSRIHDVNMOKOVFDIIGIKKS
MONERGFGLIEALKYCAPPHAGIALGLDRIVMLLTNSKNIRDVIAFPKPTTSATCLMTN
SPSTVDNLLLOLAIKHLKK"
complement(3545..4333)
/gene="znub"
complement(3545..4333)
/gene="znub"
/note="BU317"
/codon_start=1
/transl_table=11
/product="high-affinity zinc uptake system membrane
protein znub"

/protein_id="BAB13025.1"
/db_xref="GI:10038991"
/translation="MFEILFPGWLAGVLLSLTGTGPGSFTVWRMSSFEDTULSHSLL
GIALSIAFNINSFYAILIMSFATILAMLEELPVLDPVNLNISHSSLSGMVTS
LISSKKEINTITNYLFGDLISVRKNDLITTSISLILSLIFPMHSILSTINEELSO
INGINVLARITLMLTAFITIAIKFVAGALITSLIILIPATAOHFSSGPEKMWITA
IIVSILSVTGGISLVSFNTMPASPSIVLCSSFCLISLNKKHRY"
complement(4391..5107)
/gene="znuc"
complement(4391..5107)
/gene="znuc"
/note="BU318"
/codon_start=1
/transl_table=11
/product="high-affinity zinc uptake system ATP-binding
protein znuc"
/protein_id="BAB13026.1"
/db_xref="GI:10038992"
/translation="MFEFVKLRKNCVNLNHSRSLTDPDISLISNRVLTLCGPGACKS
TLVRITLGLKPSGTTIRSNLISGVYFQKHLNLTLLPTVERPKLSKRTNNTIE
EMLRKVASLRFQQLQKLSGEMORILFAKALNPNLVLDEPTQGVNGQALY
KLINERHELQCALIIVSHDLNFMVAKTDDVICLNHHICSGTPEVCNNLFEISIFG
LKVQDLATYHHNNHHHNE"
complement(6425..7867)
/gene="pyka"
complement(6425..7867)
/gene="pyka"
/EC_number="2.7.1.40"
/note="BU319"
/codon_start=1
/transl_table=11
/product="pyruvate kinase"
/protein_id="BAB13027.1"
/db_xref="GI:10038993"
/translation="MLNRLRRKIVATLGPSTDINNNEKRIIRSGVNLRFNFSHGSX
DEKSRANORAEIMRNLNFIHALDGLQPKIRSEIKNNIFLNGDPFILLDANDO
NNGNDRVGSIDYKOLPYDLKGVDELDDGRIDOLKIKSTGHEITLKYIGCLSNK
GINKLGGISADALTEKDKRDIILATEINVDYLAISFPSCSNDLKQARKLAIEFGNA
KIJAITERAEAVLNONTIEDIILASDAIMARODLGEVIGDSBLQKRLKRLRTAROL
NRIVATQMESEMINTPLPTRAEVDVNAVLDGSDVMLSAETASGEYPAETVTKM
AKICKAEKVPISNVSRIHAKFDIEEALIMASAVVANHLKGTVAITITMTESGKTA
LMSRTSGSLPIFALSKNKEITNLASLYAGVPIPYNSKNNNKANENAVILRRKGF
LCRGLVITITGDMIGKIGKINTSRILKVV"
8150..9625
/gene="zwf"
8150..9625
/gene="zwf"
/EC_number="1.1.1.49"
/note="BU320"
/codon_start=1
/transl_table=11
/product="glucose-6-phosphate 1-dehydrogenase"
/protein_id="BAB13028.1"
/db_xref="GI:10038994"
/translation="MIEETNHACDLVIFGAKGDLTKRKLLPALYKLESKKIKHYTRI
ILASGRDAMTDEYIEKIKTEVNFLEENLINDLWNLSSRIEFCNDIVBEPHPEFLK
TILKQKNKIIVYCAVPSNTLSIFIGLGNHLNVSPIVLEKPLGVCLTKSKIND
QISKYLESDQIRIDHYLGKESILNFFALFSTWCLCYFMNNKTTIDHIIYVSEEVGI
DQWVNFNMGMKMDVONHLLDITLMDQPKNTSSSIQHEKKYILRSINPIIH
NINKTVRQGYCSGVINEKKVPSYLEENAKNNSCTETVVAIKVDLNNKQSGVPYIL
RKGKRLAHRYSEIVVFKKPPTNLFNNLSELDONKLLIRLEPNPTITDFSVKAGCL
EODEYKINQLKSQFSKYSKNSIDAYERLLFEIMRGVQSLFVCHDELEAMKWIDP
IITHAMNSKNMNPOLYMSGTWGPKNSDLLLADHGRLWVYEFH"
9810..10688
/gene="htpx"
9810..10688
/gene="htpx"
/note="BU321"
/codon_start=1
/transl_table=11
/product="heat shock protein htpx"
/protein_id="BAB13029.1"
/db_xref="GI:10038995"


```

/translating="MTRIVFLTNLAVMLFSLILFNGISONTIYGLIMSGFEC
SGSLILSKWIALRSNGEITTHPNNEVESMLINVRQOSIQKTIIMPQIAVYAT
DINAFATGARRNSALIAVSGELLENNTHHEAEVYIAHEISHINGMIIITVOGVN
TEVLEIRFISQIISNMSSNRNENSTEVYLVSEFLIFLIGILASITIMFS
RHEFTADASSAMKVGKEMIALNRLKTSHEPQESDSMAFCINCKSKSPKLFASH
PSLENREALYNCEYM"
gene
11037.11246
/gene="cspc"
11037.11246
/gene="cspc"
/gene="cspc"
/note="BU322"
/codon_start=1
/translating=11
/product="cold shock-like protein cspc"
/protein_id="BAB13030.1"
/db_xref="GI:10038996"
/translating="MAKIRGOVMEKESGFGFTIPSDSKDYFVHFSSIOGKFKTL
TEGONVEFPIODGQKPAVNVFSL"
11549.13114
/gene="yoe"
11549.13114
/gene="yoe"
/note="BU323"
/codon_start=1
/translating=11
/product="hypothetical protein"
/protein_id="BAB13031.1"
/db_xref="GI:10038997"

Query Match 0.2% Score 24: DB 1: Length 293181;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 24: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3957 AAAACATGACGACGATCAATTGAAA 3980
DB 158297 AAAACATGACGACGATCAATTGAAA 158320
|||||
AC006733 58777 bp DNA linear INV 07-DEC-2001
LOCUS Caenorhabditis elegans cosmid Y32H12A, complete sequence.
DEFINITION AC006733
AC006733.2 GI:7140307
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida;
Rhabditiodea; Rhabditiidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 58777)
The C. elegans Sequencing Consortium.
Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium
Science 282 (5396), 2012-2018 (1998)
JOURNAL MEDLINE
99069613
2 (bases 1 to 58777)
Holmes, A., Elliot, G. and Cloud, J.
The sequence of C. elegans cosmid Y32H12A
JOURNAL TITLE
Unpublished
3 (bases 1 to 58777)
Waterston, R.
Direct Submission
JOURNAL TITLE
Unpublished
4 (bases 1 to 58777)
Waterston, R. H.
Direct Submission
JOURNAL TITLE
Submitted (23-FEB-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 58777)
Waterston, R. H.
Direct Submission
JOURNAL TITLE
Submitted (01-MAR-2000) Genome Sequencing Center, Washington

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (28-MAR-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
7 (bases 1 to 58777)
Waterston, R.
Direct Submission
JOURNAL TITLE
Submitted (19-JUL-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
8 (bases 1 to 58777)
Waterston, R.
Direct Submission
JOURNAL TITLE
Submitted (05-OCT-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
9 (bases 1 to 58777)
Waterston, R.
Direct Submission
JOURNAL TITLE
Submitted (07-DEC-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
On Mar 1, 2000 this sequence version replaced gi:4263205.
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1HQ, England
email: rwenematode.wustl.edu and jess@sanger.ac.uk

```

COMMENT

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

NEIGHBORING COSMID INFORMATION

This is a segment of YAC Y32H12 that was sequenced to span the gap between ZK121 and W03A5. The 5' cosmid is CELZK121, 800 base overlap; the 3' is clone CELW03A5, 200 base overlap. Actual start of this clone is at base position 6240 of CELZK121; actual end is at 32064 of CELW03A5.

NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneFINDER (P. Green and L. Hillier, ms in preparation).

FEATURES

```

source
1.58777
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="III"
/clone="Y32H12A"
complement(48..3105)
/gene="Y32H12A.2"
complement(join(48..252,503..865,1245..1626,2082..2380,
2426..2763,2813..2932,2980..3105))
/gene="Y32H12A.2"
/note="coded for by the following C. elegans cDNAs:
cm11f1, yk460e12.3, yk460e12.5, yk488b12.3, yk488b12.5,
yk492c9.3, yk492c9.5"

```

```
/codon_start=1
/product="Hypothetical protein Y32H12A.2"
/protein_id="AAF60487.1"
/db_xref="GI:7331799"
/translation="MSVNVRPVKRFVEQKKKNPEIKISEKATQLLTJPMKMSDEPKL
FMLEDEKATMGSDRIFFEEFLQDYAFILRNAVEYLDGKPTSSDTISKIARLRNTELT
HGTATLMRKREVENMDVAKLLQONVTSVEOHIOKEIDICDLDSFAGESELEVEE
OHPTEASIELANETADNEHOYLARLKPENBORRELTSLNELGRRNLOSDIRKE
VRLQGLKPKLEDLTLYAEPVEFMGAKFKDLSIEGQRKLSIQLPAPLVAIHAAAY
KEIHGEYSKILRNFLYKYLEFRLGINDTEKOSKDETEAMKRRKDETTQEKIN
KMASELLEVPMSELEFEIECSENIKISMQIYLTLEKVTAAWHINTDSKLKSPFS
IDLSGDLFENDTGDKCPMDVCEMKIEHLMKNNSAKOIGKPFKIOIESSEGT
RIGATHEHRLKVTITTRISRNRAVLGDIIRLESKKTDLNLEPEFKINSFICDES
FIASITPTSTIKTISNAAADRSTYKIEASDNGRKLTATISIPADYPRITALEGIAAP
EDSESSEVARQIEDRLNTEAHYVDVKSIKEQSLSLKHLIDS"
/genes="Y32H12A.1"
join(15954..16288,16538..16743,17240..17524,17806..17912)
/genes="Y32H12A.1"
/notes="similar to C. elegans family of proteins; see
(GI:281088)"
/codon_start=1
/product="Hypothetical protein Y32H12A.1"
/protein_id="AAF60488.1"
/db_xref="GI:7331800"
/translation="MSGATITCCRFNFIKIVKMSYIYSKMSSEHNSANTFVQCP
PTEHOILAIVLMSIKIILNSIGLNLCLMLWILISYQCLHKTFSILTYTLINDCLI
SVTLTLTKRIVOCHELTLMVFEPCPTFSTFCVLAHHLALONMLLSQILIGF
RTAMLSAOSFISMKRTITFEESYSPPIPEFTAMLLINPHIFLACYCFRHKON
NDLKRTTMSMTLRMHSSTLRKYSTKAFSLFLDLVLVLFPCQLLELPPEFTEYELG
TISSINILNPYIACCFRPYRKMEYRRLIS"
complement(18325..21785)
/genes="Y32H12A.3"
complement(join(18325..18714,20353..20604,20727..20992,
21734..21785))
/genes="Y32H12A.3"
/notes="contains similarity to pfam family PF00106 (short
chain dehydrogenase), score=124.1, E=2.6e-33, N=1; coded
for by the following C. elegans cDNAs: yk98a9.3, yk98a9.5,
yk216d4.3, yk216d4.5, yk229h3.3, yk229h3.5, yk274b4.3,
yk274b4.5, yk290g10.3, yk290g10.5, yk325b5.3, yk325b5.5,
yk325d6.3, yk325d6.5, yk374b9.3, yk428d1.3, yk428d1.5,
yk460f3.5, yk468b12.3, yk468b12.5"
/codon_start=1
/product="Hypothetical protein Y32H12A.3"
/protein_id="AAF60486.1"
/db_xref="GI:7331798"
/translation="MSLAGQIIVTGASRGIGRGIALOLGEAGATVYITGRKPEESLN
SKVGLSEATADDETTRGCGIARFVDHONMEPVKNPEVEYKEHOGOLDILVNAY
OGVATISEMKGPETEDPYWDITINNGLRNHFTCTYAAARLTARNKGLIVNVSQ
GGLRYLFNVAVGVQALDRMSADTAVELLRKNVCVVSIMPGAVRELVDKMFKDENG
KPRPEIKNAVEFANGETVEYPGRAVVSILASDRMDKTRGILITEDLGKEYGFVIDIG
LRPPMLRSVSFLIKHLGNWTTANFPTWVKLPGLVMAGNSRL"
complement(21798..21799)
/notes="S1 trans-splice site; see EST yk98a9.5"
complement(22928..25016)
/genes="Y32H12A.4"
complement(join(22928..23067,23835..23901,23954..24106,
24795..25016))
/genes="Y32H12A.4"
/notes="contains similarity to human protein phosphatase
inhibitor (SP:P41236); coded for by the following C.
elegans cDNAs: yk289g2.3, yk289g2.5, yk464f10.3,
yk464f10.5"
/codon_start=1
/product="Hypothetical protein Y32H12A.4"
/protein_id="AAF60485.2"
/db_xref="GI:14916388"
/translation="MASPSASADVDYKLTADGKICLDPAEELKRRPKKSILMKODSS
LEGKDGRAHFDEMNILATYHPTDKDYSKMKIDEPKTPYHNSDGESECEGLVGPTRR
PRRVSLGNAIDPEKVAEGLAHPESGSKISADSESDADLTTEQORAHNRDPFKKRRAH
YNEGAALKHHPLEDEDEEATAAAGAGNMEH"
complement(29721..33664)
```

```
CDS
/genes="Y32H12A.5"
complement(join(29721..29909,29971..30183,30397..31131,
31540..31907,32590..32729,32778..32881,33414..33503,
33581..33664))
/genes="Y32H12A.5"
/notes="contains similarity to schizosaccharomyces pombe
hypothetical protein (SP:I009749); coded for by the
following C. elegans cDNAs: cm1745, yk222c2.5, yk345c9.5,
yk356h10.5"
/codon_start=1
/product="Hypothetical protein Y32H12A.5"
/protein_id="AAF60489.1"
/db_xref="GI:7331801"
/translation="MEEDVESATPAESQKLLQSVNSFDERDMVVLNDDELIMEN
RNSNIIQSPLRSTVDVLRARENITIDGPIQLADTQDNQVYIDQVVOFPNSQPSQO
REETIVRSKVVVEQPASQQRDESMVYTRQNRLEPTLEGLEIDGDPVSTSKAALPFK
KGRIRAMSHPMNAKGEMKTYAVIHDNHEEGNHRRIYVYKILHPORGHEDCKHLRFV
TAGHDAEIHLDSDDDDELEVEINEEETIIPSTGEGPRAVIRKREWEAKMKTAFETL
PEMLQDNEYLTRGHRPRLPSPSCFKSIFALHETGNTIMHWGCVAFFGIGIWFLLR
PHQKQWVEKLIYSAFLCAVTCGSEFLPHYVACHSITEMAKLFSKLDYTGISLLYVC
SFPMWTFAPAYCPRQPMIITLYNGMIIITLGAAMVSLMDKFAEPKFRPRAAVFIAMGL
```

Query Match	0.2%; Score 23; DB 3; Length 58777;
Best Local Similarity	100.0%; Pred. No. 6.7;
Matches	23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 9597 TGATTTTAATTAATAAAAAAAC 9619
      |||||||
Db 27915 TGATTTTAATTAATAAAAAAAC 27937
```

Search completed: October 7, 2002, 17:59:30
Job time : 17827 secs

THIS PAGE BLANK (USPTO)


```

Db 993 caagacatctataattgcacaacattctctgcacctcctaactcttgaggaaacctcga 1052
Qy 1098 gfacctcagcttgcagctactcgaacatctggaataactcctccatgttaatgag 1157
Db 1053 gtaacttggaataatctgataagcaaacataatgtaataagcctgaagtacga 1112
Qy 1158 tcttgatgataagccttgcttaaaaggtacagaagaagaaattcttgatttaa 1212
Db 1113 aactcaccacaaaagcctcttatttaagagatacgaagagatagaacgtttaa 1167

```

RESULT 2

```

US-09-302-812-38/c
; Sequence 38, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARC) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; FILE REFERENCE: N1AD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; EARLIER FILING DATE: 1999-04-30
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 38
; LENGTH: 29793
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
; FEATURE:
US-09-302-812-38

```

Query Match 0.9%; Score 90; DB 4; Length 29793;
Best Local Similarity 50.8%; Pred. No. 5,6e-12;

Matches 309; Conservative 0; Mismatches 275; Indels 24; Gaps 3;

```

Qy 716 actcaaatgaatgatcatgtagtgcgttcaatacataatttagaattggttgaagc 775
Db 17791 ACCAATATTAATGAGCATTCGTCGTCGCCAGCTATCTTATCATTAAGTGAATG- 17731
Qy 776 cgaagacagaatgatcccaacatcaagaaactgtgatgagcgtcatglatctcac 835
Db 17732 -----TTCCGAAATGGTGAAGCGAGAAAGTCATATTAACATTGGTCGA 17687
Qy 836 tgaatttgtagatcttgctgcagtgaaagagcaagccaactgagctgaaggtgtg 895
Db 17686 CTAATCTTGTAAGATTGAGTTGCTTCAGAACGGCAATCGAAACAGAGCAACTGGCGAA 17627
Qy 896 agacttaagaaagcgtgcaacatcaacacgcagctgttacccttggcaggttattaag 955
Db 17626 AGATTTAAGAAGCTACGAAATCAATCTTCAATTATCGCTCTGGGAATGTATAGT 17567
Qy 956 aagcttagcagcgcagcagctggttgattataactacagagacagcaaacatccaga 1015
Db 17566 GCATTGGTGATGCGAATTC---TGCCCAATTCCTTATTCAGATTCGAATCAATCAGTCGA 17510
Qy 1016 attctcaaaatcatcttgaggaaatgctaaacaggttataatttcacaaattaagcca 1075
Db 17509 CTCTCTCAAGATTCTCTCGTGGAATTCAAAGACAGTTAAGTTGCTGTGATGAGACT 17450
Qy 1076 g-----ttctttgatgagactctaagctacactcagttgctgagcttaccgaacat 1129
Db 17449 GCAAGTTACATTTTGGAGACACTTGTACATTAGCATATGCAATCGACTAAAC 17390
Qy 1130 gtgagaaatactcccatgttaatgaggttccctgagatgaagcgttgcctaaagttac 1189
Db 17389 ATCAAGATCAACCGAAATTAATGAAGATCCTTAAGATGCTTACTGCGAGAAATTCGA 17330

```

```

Qy 1190 agaagaatccttgatttaagaacaattagagaatttagatcagtcgtgaaca 1249
Db 17329 GAGGAATTAATGATGCTCCGTGAGCACTGAAGCGAGAAAGACAAAGATCGAGGATGA 17270
Qy 1250 aagctcaagcaatgcttaagaagacatacagcttgctagctgaatcaaacacta 1309
Db 17269 GCAACTCAAGATTATTATGATGCTGTAACGACGAAACTGAAGATGATATGACCAATT 17210
Qy 1310 caaaga 1317
Db 17209 CAAAAAGA 17202

```

RESULT 3

```

US-09-511-477-38/c
; Sequence 38, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARC)
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREAC
; FILE REFERENCE: N1AD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 38
; LENGTH: 29793
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
; FEATURE:
US-09-511-477-38

```

Query Match 0.9%; Score 90; DB 4; Length 29793;
Best Local Similarity 50.8%; Pred. No. 5,6e-12;

Matches 309; Conservative 0; Mismatches 275; Indels 24; Gaps 3;

```

Qy 716 actcaaatgaatgatcatgtagtgcgttcaatacataatttagaattggttgaagc 775
Db 17791 ACCAATATTAATGAGCATTCGTCGTCGCCAGCTATCTTATCATTAAGTGAATG- 17731
Qy 776 cgaagacagaatgatcccaacatcaagaaactgtgatgagcgtcatglatctcac 835
Db 17732 -----TTCCGAAATGGTGAAGCGAGAAAGTCATATTAACATTGGTCGA 17687
Qy 836 tgaatttgtagatcttgctgcagtgaaagagcaagccaactgagctgaaggtgtg 895
Db 17686 CTAATCTTGTAAGATTGAGTTGCTTCAGAACGGCAATCGAAACAGAGCAACTGGCGAA 17627
Qy 896 agacttaagaaagcgtgcaacatcaacacgcagctgttacccttggcaggttattaag 955
Db 17626 AGATTTAAGAAGCTACGAAATCAATCTTCAATTATCGCTCTGGGAATGTATAGT 17567
Qy 956 aagcttagcagcgcagcagctggttgattataactacagagacagcaaacatccaga 1015
Db 17566 GCATTGGTGATGCGAATTC---TGCCCAATTCCTTATTCAGATTCGAATCAATCAGTCGA 17510
Qy 1016 attctcaaaatcatcttgaggaaatgctaaacaggttataatttcacaaattaagcca 1075
Db 17509 CTCTCTCAAGATTCTCTCGTGGAATTCAAAGACAGTTAAGTTGCTGTGATGAGACT 17450
Qy 1076 g-----ttctttgatgagactctaagctacactcagttgctgagcttaccgaacat 1129
Db 17449 GCAAGTTACATTTTGGAGACACTTGTACATTAGCATATGCAATCGACTAAAC 17390
Qy 1130 gtgagaaatactcccatgttaatgaggttccctgagatgaagcgttgcctaaagttac 1189

```

DB 17389 ATCAAAATCAACGAAATTTATATGAAGATCCCAAGCATCTTACTGCGAGAAATTCCTCA 17330
QY 1190 agaaagaaatcttggatlaaagaacaattagagaattttagagtcagtcgtgaaca 1249
DB 17329 GAAGAATTTGAATGCTCCGTGAGCACTGAAGCAGAGAAAGCAAGATTCGAGCATGGA 17270
QY 1250 aaagctcaagcaatggtctaaagaagcatalacagttctgctgaatacaacta 1309
DB 17269 GCAACTCAAGTTTATGATGCTGTAAGCAGACAAACCTTGAGATGATTTGAAGCAAT 17210
QY 1310 cacaaga 1317
DB 17209 CAAAAAGA 17202

RESULT 4
US-08-468-036-23
; Sequence 23, Application US/08468036
; Patent No. 5728806
; GENERAL INFORMATION:
; APPLICANT: Demaggio, Anthony J.
; APPLICANT: Hoekstra, Merl F.
; TITLE OF INVENTION: Materials and Methods Relating to Proteins that
; INTERACT WITH CASEIN KINASE I.
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468, 036
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/184,605
; FILING DATE: 21-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5728806and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/31784
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELE: 25-3856
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3093 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-468-036-23

Query Match 0.9%; Score 88; DB 1; Length 5093;
Best Local Similarity 49.8%; Pred. No. 6.9e-12;
Matches 313; Conservative 0; Mismatches 300; Indels 15; Gaps 3;

QY 610 taatagaaacgtgtatgttgcgtgacactgaagaacttgtaatgttccctgaacatgt 669
DB 1294 TAATCATCATCATCATGATCAAGGGGATGCAGAAATCTTTATTAATCTGCACACGAAAG 1353
QY 670 aatacagtgatcaaaaaggtgtaaaaaacagacatttgaagagactcaaatgaatga 729

DB 1354 CTGGAATTTGCTAATGCGAGGTTCGTTAAAAAAGCAAAATGCGCGCTACTAATATCAACGA 1413
QY 730 tcaatagtagtgcgtcacatacaatatttagaatgttctgaagccgagacaagaatga 789
DB 1414 TCTTTCATCAAGGCTCTCACACCGCTCTTACAAATCA--CAACAAACATAGTGGAGCANAA 1470
QY 790 tcccaaaaattcagaagaactgtgagagctgtcagatgataatccattgaatttggtaga 849
DB 1471 TAGCAAAACCATGAGCAAAAACAAAATTTGTTAAATTTGGCAAAATTTGATTTGGTGGGA 1530
QY 850 tctctgcgtgcaatgaagaagcaagccaaactggaagctgaagtggtgaacttaagaagga 909
DB 1531 TTTGGCAGGCGATGGAATAACATCAACAGATCGGCTGGCGGAATTAAGGCTCAAGAACG 1590
QY 910 ctgcaacatcaaacgcagctgttctccttggacaggttatagaagactttagcgag 969
DB 1591 TGCCCTTAATTAACAAATTCGCTGCTAACACTAGCGCGTGTATCAACGCACT-----CGT 1644
QY 970 ccaagctgtgtgattataaactacagagacagcaaacccaagatctccaaatcc 1029
DB 1645 TGATCATTTCTAACCATATACCTTACAGAGATTTAAGCTAACAAAGTTGCTACAGACTC 1704
QY 1030 atcggaggaagaatcgcataaagcgtlataatctgcacaaat-----tccgccaagctcttc 1083
DB 1705 TTTAGTGGTATGACGAAACATGATGATTTAGCAACTATATACCTCGCAAAATATTCAT 1764
QY 1084 tgatgaagcttaagtaacactcagtttgcagtaactgcacaaacatgtgaagaatctcc 1143
DB 1765 GGAAGAGACTGCAAGTACGCTTAAGATATGCAACGAGCCAAATTAAGATTAAGATCTCC 1824
QY 1144 ccatgttaatgagctcctcgtatgaatgagcgttgcctaaagaatgaagaagaatctc 1203
DB 1825 ACAAGTAATCACTCTTTATCGAAGATACATGCTTCACAGACTACATTCACAGATTTGA 1884
QY 1204 ggaattaaagaacaattagagaattta 1231
DB 1885 AAATTTAAGAAATGATTTGAATAATTCA 1912

RESULT 5
US-08-376-843-23
; Sequence 23, Application US/08376843
; Patent No. 5846764
; GENERAL INFORMATION:
; APPLICANT: Demaggio, Anthony J.
; APPLICANT: Hoekstra, Merl F.
; TITLE OF INVENTION: Materials and Methods Relating to Proteins
; INTERACT WITH CASEIN KINASE I.
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/376,843
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/184,605
; FILING DATE: 21-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5846764and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/31784
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 5093 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-376-843-23

Query Match 0.9%; Score 88; DB 2; Length 5093;
Best Local Similarity 49.8%; Pred. No. 6,9e-12;
Matches 313; Conservative 0; Mismatches 300; Indels 15; Gaps 3;

```
QY 610 taataagaagctgtatgttctgaactgaagaactgttaattgttcttgaacatgt 669
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1294 TAATTCATCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1353
QY 670 aatacagtgatcaaaaagggtgtaaaaacagacatttgagagactaaatgaatga 729
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1354 CTGAATTTCTGATGAGGTTCTGTTAAAAGGAAAGTGCCGCTACTTAATGCAACGA 1413
QY 730 tcatagtagtcgttcacatacaatattgaatgtatgttgaagcgagacagaatga 789
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1414 TCTTTATCATCAAGGCTCTACACCGCTTTTACAATCA---CACAAACATGTTGACGACAGA 1470
QY 790 tcccaaatctcagagaaactgtgagctgtcagatgatactgaactgaattgttaga 849
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1471 TACCAAGACCATGACGACAAACAAATTTGTTAAATTTGCAATTTGATTTGTTGGA 1530
QY 850 tcttctgctgagtgaaagcaagccaactgagctgaaagggtgagactaaagaaag 909
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1531 TTTGGCAGCGCAGTAAACATCACAGATCGGCTGGGAGATTAAGGCGCTCAGGAAGC 1590
QY 910 ctgcacatcaaacccagagctgtttatctcttgagacagtgatataagaagcttagcgacgg 969
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1591 TGGCCATATTAACAAATCGCTCTTACACTAGCGCTGTATCAACGCACT-----CGT 1644
QY 970 ccaagctgtgattataactacagagacagcaaacctcaccagaacttcccaaatc 1029
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1645 TGATCTATCTTAACCATATATCTTACAGAGAAATCTAAGTACACAGATTGCTACAGATC 1704
QY 1030 attggaggaagaattgtaaaacggtatataattgcaaat-----tacgcagtttctt 1083
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1705 TTTAGTGTATGACGAAACATGATGATATGCAACTATATCACCTCGGAAATATATCCAT 1764
QY 1084 tgatgagactctaaagtacactcagtttgccagttactcacaacatgtgagaatactcc 1143
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1765 GGAAGAGACTGCAACTAGCTGATGATATGCAACGACCAACCAATCAATTAAGAACTACTCC 1824
QY 1144 ccatgtaataaggtctcctgtagtgaagcgttctaaaaagtgacagaagaatcct 1203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1825 ACAAGTAAATAGCTTTATGGAAGATACATGCTCAAAAGACTACATCAAGAGATTGA 1884
QY 1204 ggaattaaagaacaattagaaatta 1231
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1885 AAATTTAAGAAATGATTGAAAAATTC 1912
```

RESULT 6
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 2213-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZpT-Fls
US-08-232-463-14

Query Match 0.9%; Score 85.2; DB 1; Length 7218;
Best Local Similarity 4.5%; Pred. No. 4.1e-11;
Matches 18; Conservative 245; Mismatches 133; Indels 0; Gaps 0;

```
QY 5271 taaagaagaattgattgggttgaatttaaaatgaagcgagcaaaagaccacta 5330
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1459 TTAAGAGATAGAGAATTTGCTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1400
QY 5331 aagaacatgtctgtaaatgaatlaaagaacttgagcaagccagacacagactcaat 5390
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1399 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1340
QY 5391 gtgagtagaagaacttagaagaacttaagaagctcagcgctggaactta 5450
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1339 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1280
QY 5451 aagaatctgaacaaaagtaattaaactgaacaaagaaatgaaatgtttagtcggaa 5510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1279 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1220
QY 5511 tggaggaattgaaaatagccagagactgtlaattgtcagagagacagctgcaagc 5570
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1219 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1160
QY 5571 acctaagagagtggtgaaatgcatctgaactcagaagatcagaagaagctcaag 5630
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1159 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1100
QY 5631 aagcatcagcagcagaagaataaagctcaggaac 5666
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1099 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRATC 1064
```

RESULT 7

RESULT 9
US-09-572-191-5
; Sequence 5, Application US/09572191
; Patent No. 6355466
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6355466el motor proteins and methods for
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/572.191
; CURRENT FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1230
; TYPE: DNA
; ORGANISM: Human
US-09-572-191-5

Query Match 0.9%; Score 82; DB 4; Length 1230;
Best Local Similarity 57.0%; Pred. No. 9.8e-11;
Matches 195; Conservative 0; Mismatches 135; Indels 12; Gaps 2;
QY 830 tctcaattgaattgtagatctgtctgagtgaaagagcaagcacaactgagctgaa 889
DB 775 tccctactcaacctggtgatttagcagatcgtgaagaaagaaatcccatgcaaa 834
QY 890 ggtgtgagacttaagaaagctgcaacatacaccgacgtgtttatctcttgacaggt 949
DB 835 gggatgagacttgaaagaaagaaagaaatcgaatacgtatgagctggtgccaagtg 894
QY 950 attaagaagcttagcagcagcagcgtggtgtgattataa-----ctacagaacacg 1003
DB 895 attacagcacttgctgagctggtgtaatgaaacagacagactgttctacagagactcc 954
QY 1004 aaactcaccagaattctccaaatcattcattgaggaatgctaaacggtataattgc 1063
DB 955 aaactcactcttactacgagatctcccttgaggttaagtccaacacgacataattgca 1014
QY 1064 acaattacgcagttc-----ttttagtgagacttaagtaactcagttgcaggt 1117
DB 1015 aatgttcacccgtgacccagctgttttgaggaaaccctatacaacttaacttgcctcaa 1074
QY 1118 actgccaacatgtgagaataactcccatgttaatgagtc 1159
DB 1075 agagccaagctgattataaacaagcagtagtaaatgaaagac 1116

RESULT 10
US-09-572-191-1
; Sequence 1, Application US/09572191
; Patent No. 6355466
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6355466el motor proteins and methods for
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/572.191
; CURRENT FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4757
; TYPE: DNA
; ORGANISM: Human
US-09-572-191-1

Query Match 0.9%; Score 82; DB 4; Length 4757;
Best Local Similarity 57.0%; Pred. No. 2e-10;
Matches 195; Conservative 0; Mismatches 135; Indels 12; Gaps 2;
QY 830 tctcaattgaattgtagatctgtctgagtgaaagagcaagcacaactgagctgaa 889
DB 775 tccctactcaacctggtgatttagcagatcgtgaagaaagaaatcccatgcaaa 834
QY 890 ggtgtgagacttaagaaagctgcaacatacaccgacgtgtttatctcttgacaggt 949
DB 835 gggatgagacttgaaagaaagaaagaaatcgaatacgtatgagctggtgccaagtg 894
QY 950 attaagaagcttagcagcagcagcgtgtgtgattataa-----ctacagaacacg 1003
DB 895 attacagcacttgctgagctggtgtaatgaaacagacagactgttctacagagactcc 954
QY 1004 aaactcaccagaattctccaaatcattcattgaggaatgctaaacggtataattgc 1063
DB 955 aaactcactcttactacgagatctcccttgaggttaagtccaacacgacataattgca 1014
QY 1064 acaattacgcagttc-----ttttagtgagacttaagtaactcagttgcaggt 1117
DB 1015 aatgttcacccgtgacccagctgttttgaggaaaccctatacaacttaacttgcctcaa 1074
QY 1118 actgccaacatgtgagaataactcccatgttaatgagtc 1159
DB 1075 agagccaagctgattataaacaagcagtagtaaatgaaagac 1116

RESULT 11
US-08-480-552-18
; Sequence 18, Application US/08480552
; Patent No. 5665350
; GENERAL INFORMATION:
; APPLICANT: Gudkov, Andrei
; APPLICANT: Roninson, Igor B.
; TITLE OF INVENTION: Genes And Genetic Elements Associated
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allegretti & Witcoff, Ltd.
; STREET: 75 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480.552
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/033,086
; FILING DATE: 09 MAR 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Keown, Wayne A.
; REGISTRATION NUMBER: 33,923
; REFERENCE/DOCKET NUMBER: 93,354
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/345-9100
; TELEFAX: 617/345-9111
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2389 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA

US-08-480-552-18
; ANTI-SENSE: YES
; HYPOTHETICAL: NO

Query Match	0.8%;	Score 79.8;	DB 1;	Length 2389;
Best Local Similarity	49.9%;	Pred. No. 4.9e-10;		
Matches 430;	Conservative 0;	Mismatches 377;	Indels 54;	Gaps 7;

[illegible]

```

? Patent No: 5866327
?
? GENERAL INFORMATION:
?
? APPLICANT: Gudkov, Andrei
?
? APPLICANT: Roninson, Igor B.
?
? TITLE OF INVENTION: Association of Kinesin with Sensitivity
?
? TITLE OF INVENTION: To Chemotherapeutic Drugs
?
? NUMBER OF SEQUENCES: 8
?
? CORRESPONDENCE ADDRESS:
?
? ADDRESSEE: Allegretti & Witcoff, Ltd.
?
? STREET: 75 State Street
?
? CITY: Boston
?
? STATE: Massachusetts
?
? COUNTRY: U.S.A.
?
? ZIP: 02109
?
? COMPUTER READABLE FORM:
?
? MEDIUM TYPE: Floppy disk
?
? COMPUTER: IBM PC compatible
?
? OPERATING SYSTEM: PC-DOS/MS-DOS
?
? SOFTWARE: Patent in Release #1.0, Version #1.25
?
? CURRENT APPLICATION DATA:
?
? APPLICATION NUMBER: US/08/486,382
?
? FILING DATE:
?
? CLASSIFICATION: 435
?
? PRIOR APPLICATION DATA:
?
? APPLICATION NUMBER: US 08/177,571
?
? FILING DATE: 05 JAN 1994
?
? ATTORNEY/AGENT INFORMATION:
?
? NAME: No. 5866327nan, Kevin E.
?
? REGISTRATION NUMBER: 35,303
?
? REFERENCE/DOCKET NUMBER: 93,354-A
?
? TELECOMMUNICATION INFORMATION:
?
? TELEPHONE: 312/715-1000
?
? TELEFAX: 312/715-1234
?
? INFORMATION FOR SEQ ID NO: 4:
?
? SEQUENCE CHARACTERISTICS:
?
? LENGTH: 2389 base pairs
?
? TYPE: nucleic acid
?
? STRANDEDNESS: single
?
? TOPOLOGY: linear
?
? MOLECULE TYPE: cdna
?
? HYPOTHETICAL: NO
?
? US-08-486-382-4

```

Query Match	0.8%	Score 79.8	DB 2:	Length 2369
Best Local Similarity	49.9%	Pred. No. 4.9e10		
Matches 430	Conservative	0	Mismatches 377	Indels 54
			Gaps	7
QY	393	gacagacatcttcacgacgaagcgtcacacagatgagacc-----ccaaatcat	443	
Db	2	GACAAACATCATCTGTGGAAAGCCACAGATGGAGGTAACCTCATGATCTCAGAGCA	61	
QY	444	tgggcataatacccccaagccatcacggaagttttaaataatccagaagat---accga	500	
Db	62	TGGGAATTAATTCAGAAATAGTCAGAGATATTTTAAATTAATTAATTCACATGATGATA	121	
QY	501	acagagagttctcttaagagttctcttaaggagatttacaagaacgtglgaagacg	560	
Db	122	ATTTCGAATTCATATTAAGTTTCATATTTTGAATAATTAATTTGGATTAAGATAAGGGACT	181	
QY	561	tactctgtgatgcagaaagaagacccttggaaatccgagagattttaatgaabaacg	620	
Db	182	TGTTAGATGTT-----TCAAAAGACTTAACCTTTTCAGTCCATGAAAGACAAAACGGTGTTC	235	
QY	621	tgtatgttcgtgacctgacgtcgaagaacttctaagttctccggaacgttaatacagtga	680	
Db	236	CCTAATGTAAGGGGCGCACAGAACTTTTCGTGTGATGCCAGATGAATCTCATGTGATGCCA	295	
QY	681	tcaaaaagggtgaaaaaaacagacgttaatggagagctcaaaatgaatgaatgaatgtagtc	740	
Db	296	TAGATGAAGGGGAAATCCAAACAGATGTGTCCGAGTTACAAATATATGATATGACATAGCTCTA	355	
QY	741	gtccacatacaaatlttaagaatgattgttgaagccgagacagaatgatccccaatt	800	

```

Db 356 GGAGCCACGACATATTTCTTATTAATGTAACAA-----AGACATATACAAACGG 406
QY 801 cagagaactgtagagcgtgcatgtagtctcaactgaattggtagactgtcgcgga 860
Db 407 AACAGAACTCAGTGA-----AAGCTTTATCTGGTGTGATTAAGTCGCA 451
QY 861 gtgaagaagcaagcaaacctggaagctgtagaagctgtagaagcaagcgtgcaacatca 920
Db 452 GTGGAAGGTTAGTAAGACTGCGGCTGAAGTGCTGTGATGAAGCAACATCA 511
QY 921 accgcaactgtttatcctctggacaggtatattaagaactttagcagcgccagctggtg 980
Db 512 AGAAGTACTTTCTGCACCTTGGAAATGTCTCTTCTGCGCAGAGGCG-----AGTA 565
QY 981 gattataaactacagagcaagcaaacctccagaaattcccaaatatttggagagga 1040
Db 566 CCTATGTTCTTATTCGAGATAGTAATAATGACCGAATTTCTTCAGATTTCATTAAGTGCA 625
QY 1041 atgctaaagcgttataatttgcacaattacgcagttcttcttgaagactctaagt- 1099
Db 626 ACTGTAGGACCACTATTGTCTATATGCTGCTCTCATCATCATATCAATGAGTCTGAGACA 685
QY 1100 -----acacttcagtttgcagctgcaacaacatgtagaatactcccaattgaatg 1154
Db 686 AGTCAACACTCTCTTTGGTCAAAAGGCCAAACAAATTAAGAACACAGTCTGTGCAATG 745
QY 1155 aggtcctgtagatgtagaagcgttcttaaaaaggtacagaagaagaatcttgaattaaaga 1214
Db 746 TAGAGTTAACTGACAGAGCAGTGAAGAAAGAACATTAAGAAAGAAAAGAAAATAAAGA 805
QY 1215 aacaattagagaatttagagt 1235
Db 806 CTCTACGGAACACTATTCAGT 826

RESULT 13
US-09-235-546-4
; Sequence 4, Application US/09235546
; Patent No. 6043340
; GENERAL INFORMATION:
; APPLICANT: Gudkov, Andrei
; APPLICANT: Roninson, Igor B.
; TITLE OF INVENTION: Association of Kinesin with Sensitivity
; TITLE OF INVENTION: No Chemotherapeutic Drugs
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allegretti & Wilcoff, Ltd.
; STREET: 75 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/235,546
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,382
; FILING DATE:
; APPLICATION NUMBER: US/08/177,571
; FILING DATE: 05 JAN 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6043340nan, Kevin E.
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 93,354-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/715-1000

```

```

; TELEFAX: 312/715-1234
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2389 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEetical: NO
; US-09-235-546-4

Query Match      0.84; Score 79.8; DB 3; Length 2389;
Best Local Similarity 49.98; Pred. No. 4, 9e-10;
Matches 430; Conservative 0; Mismatches 377; Indels 54; Gaps 7;

QY 393 gacagacatcttcagcaagaagcgtacacaattgttggagaca-----ccaaattcat 443
Db 2 GACAACATCATCTGGGAGAGAGCCACACGATGAGGGTTAACTTCATGATCGAAGGCA 61
QY 444 tgggcaataaaccccaagcacaagaagttttaaatattcaggaat---accga 500
Db 62 TGGGAATTATTCAGAAAGATAGTGAAGATATTTTATTAATTAATTTACCTCATGATGATA 121
QY 501 acagaaggttctctcctaagagttcttataatgagtttacaatgaactgtgaaagacc 560
Db 122 ATTTGGAAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 181
QY 561 tactgtgtgtagacagaagaagaagcccttggaaattcgcgaagatttaataagaagc 620
Db 182 TGTATGATGTT-----TCAAGACTAACCTTTCACCTCATGATGATGATGATGATGATGAT 235
QY 621 tgtatgtgtcgtacctgactgtagaagaactgttaattgttccgtgaacatgtaacagtaga 680
Db 236 CTATGTAAAGGGGTGACAGAACGTTTGTGTGTAGTCCAGATGAAGCATGATGATGATGATGAT 295
QY 681 tcaaaaaggtgaaaaaaacagacattatgtagagagactaaatgtagatcatagtagtc 740
Db 296 TAGATGAAGGGAATTCACACAGATGATGCGAGTTTCAAAATTAATGAATGAATGAATGATGAT 355
QY 741 gtccacatcaaatatttagaatgtagttagaagccgagacagaagaatgatcccaaat 800
Db 356 GGAGCCACAGCAATTTCTTATTAATGTAACAA-----ACAGAAATACAAACGG 406
QY 801 cagagaagcgtgtagagctgtagcagttatctcaactgaatttggtagactgtcgga 860
Db 407 AACAGAACTCAGTGA-----AAGCTTTATCTGGTGTGATTAAGTCGCA 451
QY 861 gtgaagaagcaagcaaacctggaagcgtgtagaagctgtagaagcaagcgtgcaacatca 920
Db 452 GTGGAAGGTTAGTAAGACTGCGGCTGAAGTGCTGTGATGAAGCAACATCA 511
QY 921 accgcaactgtttatcctctggacaggttattaagaagcttagcagcgccagctggtg 980
Db 512 AGAAGTACTTTCTGCACCTTGGAAATGTCTCTTCTGCGCAGAGGCG-----AGTA 565
QY 981 gattataaactacagagcaagcaaacctccagaaattcccaaatatttggagagga 1040
Db 566 CCTATGTTCTTATTCGAGATAGTAATAATGACCGAATTTCTTCAGATTTCATTAAGTGCA 625
QY 1041 atgctaaagcgttataatttgcacaattacgcagtttcttgaagactctaagt- 1099
Db 626 ACTGTAGGACCACTATTGTCTATATGCTGCTCTCATCATCATATCAATGAGTCTGAGACA 685
QY 1100 -----acacttcagtttgcagctgcaacaacatgtagaatactcccaattgaatg 1154
Db 686 AGTCAACACTCTCTTTGGTCAAAAGGCCAAACAAATTAAGAACACAGTCTGTGCAATG 745
QY 1155 aggtcctgtagatgtagaagcgttcttaaaaaggtacagaagaagaatcttgaattaaaga 1214
Db 746 TAGAGTTAACTGACAGAGCAGTGAAGAAAGAACATTAAGAAAGAAAAGAAAATAAAGA 805
QY 1215 aacaattagagaatttagagt 1235

```



```

;
; FILING DATE: 09-SEP-1997
; ATTORNEY/AGENT INFORMATION:
;   NAME: Keown, Wayne A.
;   REGISTRATION NUMBER: 33,923
;   REFERENCE/DOCKET NUMBER: 93,354
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 617/345-9110
;   TELEFAX: 617/345-9111
; INFORMATION FOR SEQ ID NO: 18:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 2389 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;     MOLECULE TYPE: cDNA
;     HYPOTHEetical: NO
;     ANTI-SENSE: YES
;   SEQUENCE DESCRIPTION: SEQ ID NO: 18:
;
US-09-568-315-18
```

Query Match 0.8%; Score 79.8; DB 4; Length 2389;

Best Local Similarity 49.9%; Pred. No. 4.9e-10;

Matches 430; Conservative 0; Mismatches 377; Indels 54; Gaps 7;

```

QY 393. gacagacatcttcagcaagaacgtacacatgatggaaca-----ccaattcat 443
DB 2 GACAAACATCATCTGGGAGACCACACAGATGAGGGTAACTTCATGATCCAGAGGCA 61
QY 444 tgggcatatcccccaagccatagagaagtttaaatattcagagaat---accga 500
DB 62 TGGGATATTATTCAGATAGTATGCAAGATATTTTAAATATATTTACTCCATGATGAAA 121
QY 501 acaagagattcttcttaagattcttataatgagatttacaatgaaactgtgaagacc 560
DB 122 ATTGGAATTTCAATTAAGGTTTCATATTTTGAATATATTTGATTAAGATTAAGGACT 181
QY 561 taactgtgtgagacagaagaagcccttggaattcgagagattttaagaaacg 620
DB 182 TGTGATGATGTT-----TCAAGACTTAACCTTTCAGTCCATGAGACAAAACCGTTC 235
QY 621 tgaatgtgtcactgactgagaactgttaatggttcctgaacatgtaatacagltga 680
DB 236 CCTATGTAAGGGGTGCACAGACCTTCGTGTAGTCCAGATGAGTCAATGATGATACCA 295
QY 681 tcaaaaagggtgaaaaaaacagacattatgagagactaaatgaaatgataatgtagtc 740
DB 296 TAGATGAAGGGAAATCCAAACAGAGATGTCGCAATTACAAATATGAATGAACATGACTCTA 355
QY 741 gtacacatacaatattgaatgattgttgaagaagccgagacagaatgatccacaatt 800
DB 356 GGAGCCACACGCAATTTCTTATTATGTAACA-----ACAGAAATACACAAACGG 406
QY 801 cagagaactgtgtgagactgtcatgtatctcacttgaatttggtagactgtctgca 860
DB 407 AACAGAACTCAGTGA-----AAGCTTATCTGTTGAGTGAAGCTGCA 451
QY 861 gtgaagaagcaagccaaactgtgaagctgtgaagctgaagactgaagaagctgacatca 920
DB 452 GTGAGAGGGTTAGTAAGACTGGGGCTGAAGGTGCTGTGATGATGAAGTAAAGACATCA 511
QY 921 accgacagctgttattccttgagcaaggttattlaagaagcttagcagcgccagctggtg 980
DB 512 AGAAGTCACTTTCTGCATTTGCAAAATGTCATTTCTGTTGGCAGAGGSC-----AGTA 565
QY 981 gattataactacagagaacagcaaacacacagaattctccaagaattcatttggagagaa 1040
DB 566 CTTATGTCTCTTATCGATAGTAAGTAAGTAAATGACGAAATTTCTCAAGATTCATTAGTGGA 625
QY 1041 atgcttaaacggttataatttgacaattacagccagttcctttagagactctaagt- 1099
DB 626 ACTGTAGAGACCACTATTGCAATATGCTGCTCTCATCATCATATACATATGAGTCTGAGACAA 685
```

```

QY 1100 -----acacttcagtttgcagtaactgccaacatgttgagaataactcccaatgtaatg 1154
DB 686 ACTCAACACTCTCTTTGTTGTCAAAGGCCAAACATTTAAGAACACAGTCTGTGCAATG 745
QY 1155 aggtcctgagatgaagcgttgcctaaaaaggttacagaagaaggaatcttgatttaaga 1214
DB 746 TAGAGTTAACTGACAGACAGTGGAAAAAGAGTATGAAAAAGAAAAAANTATAGA 805
QY 1215 aacaaattagagaatttagagt 1235
DB 806 CTCTACGGAACACTATTTCAGT 826
```

Search completed: October 7, 2002, 09:06:54
Job time: 486 sec

THIS PAGE RI ANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 7, 2002, 08:58:48 ; Search time 9778.42 Seconds
(without alignments)
13286.573 Million cell updates/sec

Title: US-09-150-867-2

Perfect score: 9626
Sequence: 1 gaattcggcggtcgatag.....ttaaaaaaaaaacgaattc 9626

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estha:*
2: em_esthum:*
3: em_estin:*
4: em_estinu:*
5: em_estlov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_estl2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	600.8	6.2	636	10	BM262378 daq40c12.
2	595.8	6.2	681	9	AL638330 AL638330
3	541	5.6	554	9	AM641354 AM641354
4	531.6	5.5	669	10	BJ069784 BJ069784
5	526.4	5.5	583	10	BM262071 BM262071
6	511.4	5.3	545	10	BM262071 daq40c12.
7	453.4	4.7	609	10	BM262071 daq40c12.
8	433.6	4.5	670	9	AL659719 AL659719
9	424.4	4.4	426	10	BI349745 dae63c09.
10	423.8	4.4	457	9	AM636063 BI426c09.
11	404.4	4.2	612	10	BJ074200 BJ074200
12	378.4	3.9	521	9	AL644957 AL644957
13	371.6	3.9	408	10	BM262071 daq40c12.
14	369.8	3.8	646	10	BM262071 daq40c12.
15	318	3.3	798	9	BM262071 daq40c12.
16	313.8	3.3	430	10	BM262071 daq40c12.
17	273	2.8	596	9	AL644957 AL644957

18	258.4	2.7	525	9	AA907657
19	254.2	2.6	533	10	UB3037
20	251.8	2.6	430	9	AA964355
21	241	2.5	639	9	AA623883
22	238.4	2.5	546	9	AA811002
23	235.4	2.4	939	10	BE573123
24	234.8	2.4	613	10	BC081419
25	232.6	2.4	1043	10	BE573608
26	221.6	2.3	547	9	AI505210
27	216	2.2	443	9	AA983741
28	203.4	2.1	806	9	AU131640
29	193	2.0	614	10	BJ088279
30	189.2	2.0	429	9	AM449147
31	176.2	1.8	409	9	AA721157
32	165.4	1.7	502	10	BE636970
33	165	1.7	376	9	AA287391
34	164.6	1.7	743	10	BI925210
35	161.6	1.7	552	10	BE721105
36	152.6	1.6	757	10	BI933039
37	149.6	1.6	724	10	BI931144
38	149.2	1.5	648	10	BE305901
39	147.8	1.5	533	10	BU086865
40	147.4	1.5	567	10	BM225822
41	147	1.5	557	10	BM228285
42	146.8	1.5	559	10	BM227569
43	146.8	1.5	653	9	AW735873
44	146.8	1.5	657	11	BC020890
45	143.2	1.5	657	10	BM229569

ALIGNMENTS

RESULT 1
LOCUS BM262378 636 bp mRNA linear EST 18-DEC-2001
DEFINITION daq40c12.Y3 Blackbear/Soares normalized Xenopus egg library
Xenopus laevis cDNA clone IMAGE:4783871 5' similar to: TR:O42263
O42263 KINESIN-RELATED PROTEIN. ; mRNA sequence.

ACCESSION BM262378
VERSION BM262378.1 GI:17925418
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis

REFERENCE
AUTHORS Eukariyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 636)
Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D.,
Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y., Person,
B., Gibbons, M., Harvey, N., Rilter, E., Jackson, Y., McCann, R.,
Waterston, R. and Wilson, R.
Washu Xenopus EST project, 1999
Unpublished (1999)
Other ESTs: daq40c12.x3

TITLE
JOURNAL
COMMENT
Contact: Sandy Clifton, Ph.D.
Washu Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estelw@wustl.edu
Library constructed by Bento Soares and M. Patricia Bonaldo
(University of Iowa). DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LN01 at: info@imgc.llnl.gov
Seq primer: -40RP from Gibco
High quality sequence stop: 493.
Location/Qualifiers
1..636
/organism="Xenopus laevis"
/db_xref="taxon:8335"

FEATURES

source


```
/clone="IMAGE:4783871"
/clone_lib="Blackshear/Soares normalized Xenopus egg
library"
/sex="female"
/tissue_type="unfertilized egg"
/cell_type="unfertilized egg"
/dev_stage="unfertilized egg"
/lab_host="DH10B"
/Note="Vector: pT73-Pac; Site_1: EcoRI; Site_2: NotI;
PolyA-selected mRNA was prepared from unfertilized Xenopus
laevis eggs. The library was constructed in the vector
pT73-Pac as described in Bonaldo, M.F., Lennon, G. and
Soares, M.B. 'Normalization and subtraction: two
approaches to facilitate gene discovery', Genome Research
5:791-806, 1996. The first strand synthesis used a
NotI-dT18 primer; double stranded cDNAs were ligated to
EcoRI adaptors, digested with NotI, and directionally
cloned into the NotI and EcoRI-digested pT73-Pac vector.
The library contained approximately 7.2 x 105
recombinants, with average insert sizes of 1-1.5 kb."
```

BASE COUNT 185 a 126 c 125 g 200 t

ORIGIN

Query Match 6.2%; Score 600.8; DB 10; Length 636;
Best Local Similarity 99.0%; Pred. No. 1.8e-104;
Matches 616; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

```
QY 8794 gcaaaaagtcacaggaataatcccccataaggaataattatgtatgaatcaaa 8853
DB 15 GCAGAAAGTACAGAGAAATCTAATCCCTTAAAGCAATATTGATGTAGAGTCAAA 74
QY 8854 atcagatgcttaactgcatctcaatgtttttatataatttaagctgtgtatcttca 8913
DB 75 ATCGATGCTTACTGCTGCATCAGTCTTTTATATTTTAAAGCTGGATGTTTCA 134
QY 8914 gctcaacacagagagagacatgacaaaagtcagctggaactgtgtatgaagcaaa 8973
DB 135 GCTCAACACAGACAGAGAGCAATGACAAAGTGCAGGCTGAGACTGTGTATGAAGCAA 194
QY 8974 aaaaagaacgacactgaaatgtaaaacatcctagatccctgtgaactctgactctcttc 9033
DB 195 AAAAGAACGACCTGAAATGTAACATCCTAGATCCCTGTACATCTGACCTCTCTGC 254
QY 9034 ctgcaaaagagactgtctactctgctctctgtgaagaagaacacagaacatgcacatgc 9093
DB 255 CTGCAAGAGACTGCTACTGCTGCTCTGTGAAGAAACACTAAGAACGCCATGTC 314
QY 9094 tgcataaagagatcctcactcgtgaagaagaagctgtctcttaagtaaatcactgtgtggc 9153
DB 315 TCCATTAAGAGAGTCCACTGGAAGCAAAAGTGTCTTATAGTAATCACCTGGTGGGC 374
QY 9154 gagggttaacgtcttaataaagtgcaatacgtctcagctctatattatattatgt 9213
DB 375 GAGTGGTTACGCTTTAAATTAAGTGAATTCGCTCAGCTCTATTATATATGTTAAGT 434
QY 9214 ctgtatattgttacactttaagtccttgactcatatttggctcactgtatgttct 9273
DB 435 CTGTCTATTGTTTACACTTTTAAAGTCCCTGACTCATATTGAGCTCATGTTGTTCT 494
QY 9274 ttgtgttgccgaacac-----acacactgtgtaaaaaatgacattggcaatgatttca 9329
DB 495 TTGTGTTTGGCCACACATACACACTGCTGTAGAAAATACATTGCACTGATATTGTTTA 554
QY 9330 ctgactggtctctctggtggccatcatgcaaaagcaccattgtgtgcaatgattttcac 9389
DB 555 CTGACTGCTCTCTGCGGGCCATCATGCAAAAGCACCTTATGTTGCAATGTTTTCAC 614
QY 9390 tacttatattatattgtcactt 9411
DB 615 TACTTATTATTATGTCGACTT 636
```

RESULT 2
AL638330 681 bp mRNA linear EST 12-DEC-2001
LOCUS AL638330 XGC-neurula silurana tropicalis cDNA clone TNeu018h15 5',
DEFINITION mRNA sequence.
ACCESSION AL638330.1 GI:16790309
VERSION AL638330.1
KEYWORDS EST.
SOURCE Western clawed frog.
ORGANISM Silurana tropicalis
Euryptera; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
Xenopodinae; Silurana.
REFERENCE 1 (bases 1 to 681)
AUTHORS Huckle, E., Taylor, R., Ashurst, J. L., Zorn, A. M. and Rogers, J.
TITLE Sanger Xenopus tropicalis EST project 2001 (10-2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Huckle E
Sanger Centre
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS SEQUENCE ID: TNeu018h15.sp6
Sequencing primer: Sp6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
Location/Qualifiers
1. .681
/organism="Silurana tropicalis"
/db_xref="taxon:8364"
/clone="TNeu018h15"
/clone_lib="XGC-neurula"
/dev_stage="neurula"
/lab_host="Escherichia coli DH10B"
/note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from 5ug of poly A+ RNA from neurula.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end."

FEATURES
source

BASE COUNT 224 a 129 c 158 g 169 t 1 others

ORIGIN

Query Match 6.2%; Score 595.8; DB 9; Length 681;
Best Local Similarity 93.5%; Pred. No. 1.6e-103;
Matches 632; Conservative 0; Mismatches 43; Indels 1; Gaps 1;

```
QY 100 gggagcggtgtcggtaccagatttcacactaatcgttcacaaaatgtccgaggagatgc 159
DB 7 GGGAGCGGTGTGCGGTACCGATTTCGACATATGCGACTC-AAATGCCAGGAAGCGC 65
QY 160 agttaagtggtgtgaggggttcggccgttatcagagagaagaaggagataagccaa 219
DB 66 AGTAAAGTGTGTGTGAGGTCGCGCGCTTATACAGAGAGCAAGGGGATCAATCCAC 125
QY 220 ccttgaaatggaaggcttgaaacaacacacattcccaagtgtgatggaagaagcttcaa 279
DB 126 CCTCTATWGAAGCGCTGGAAGCAACACATTCCAGTGTAAGGGAGCAAAATCTTTCA 185
QY 280 tttagacgtgtattatattcactcaagaatacaagaatcaaatlaccaagaatagcagt 339
DB 186 TTTTGATGCGGTATTTAATTCACGAAATCAACAAAGCAAGTATTACCAAGAAATATACAGT 245
QY 340 acctatatacgtatcaactcttgcaaggagataatgacacataattgtcagtaggaagac 399
DB 246 ACCATATATACAGTCACTTTGCAAGGATATTAATGGACAAATTTTGGCTTATGACAGAC 305
QY 400 atttcaggaacagtcacaaatgalyggaacaaccaaatltcatlttggcataatccca 459
DB 306 ATCTTCAGCAAGACCTACACGATGATGGAGAACACCAAAATTTTGGGATATATACCCA 365
QY 460 agccatacaggaagttttaaatattcagagagataccgaaagaagagttcttctaag 519
DB 366 AGCCATACAGGAAGTGTTAATAATTCACAGAGATACCCAAACAGAGAGTTCCTCTAAG 425
```

0y 520 agttccttatatggaattacatgaacacgtgaaagaccctactgtgtgatgcagaag 579
|||||
Db 426 acTTTCTATATGCAATTTACATGAACCGTGAAGACCTCTCTGTGATGACAGAAA 485
0y 580 aagaagacccttggaatcgcgagatttaataagaacgtgtatgtctgacctgac 639
486 AAGAAGACCCCTTGGAATTCGCGAGGATATCAATGAAGACGTATGTGTGCTGACTTAA 545
0y 640 tgaagaactgtgaatgtgtccggaacatgtaatacagtgatcaaaaaggtgaaaaa 699
546 AGAAGACCTTGGAATGTCTCCGACCATGTAATTCAGTGATCAAAAAGGTCGAAAAA 605
0y 700 cagacattatgagagactaaatgaatgatcatagtagtgcctcaacataattttg 759
606 CAGGATTTATGAGAGACTAAATGATGATCATAGTACCGCTCCCATACATATNTCG 665
0y 760 aatgattgtgaaagc 775
666 AATGATTGTGAAAGC 681
Db
RESULT 3
AM641354 554 bp mRNA linear EST 26-APR-2001
LOCUS
DEFINITION laevis cDNA clone PBX0105E04 5', mRNA sequence.
ACCESSION AM641354
VERSION AM641354.1 GI:7398613
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 554)
Blackshear, P.J., Lai, W.S., Thorn, J.M., Kennington, E.A., Staffa, N.G.,
Moore, D.T., Bouffard, G.G., Beckstrom-Sternberg, S.M., Touchman
J.M., Bonaldo, M.F. and Soares, M.B.
The NIHES Xenopus maternal EST project: interim analysis of the
first 13,879 ESTs from unfertilized eggs
Gene 267 (1), 71-87 (2001)
21211403
Contact: Perry J. Blackshear
Office of Clinical Research and Laboratory of Signal Transduction
National Institute of Environmental Health Sciences
A2-05 NIHES, 101 Alexander Drive, Research Triangle Park, NC 27709,
USA
Tel: 919 541-4899
Fax: 919 541-4571
Email: black009@niehs.nih.gov
Clone is available through Research Genetics, Inc., 2130 Memorial
Parkway, Huntsville, AL 35901
Phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email
cdna@resgen.com
DNA sequencing and analyses performed by National Institutes of
Health Intramural Sequencing Center (NISC).
PCR Primers
FORWARD: TGTAAACGACGCGCCACT
BACKWARD: CAGGAACAGCATGACG
Plate: 0105 row: E column: 04
Seq primer: 77 primer.
Location/Qualifiers
1. 554
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="PBX0105E04"
/clone_lib="Blackshear/soares normalized Xenopus egg
library"
/sex="female"
/tissue_type="unfertilized egg"
/cell_type="unfertilized egg"
/dev_stage="unfertilized egg"

/lab_host="DH10B"
/note="Vector: pT773-Pac; Site.1: EcoRI; Site.2: NotI;
PolyA-selected mRNA was prepared from unfertilized Xenopus
laevis eggs. The library was constructed in the vector
pT773-Pac as described in Bonaldo, M.F., Lennon, G. and
Soares, M.B. 'Normalization and subtraction', Genome Research
approaches to facilitate gene discovery', Genome Research
6:791-806, 1996. The first strand synthesis used a
NotI-dT18 primer; double stranded cDNAs were ligated to
EcoRI adapters, digested with NotI, and directionally
cloned into the NotI and EcoRI-digested pT773-Pac vector.
The library contained approximately 7.2 x 10⁵
recombinants, with average insert sizes of 1-1.5 kb."

BASE COUNT 177 a 107 c 129 g 141 t
ORIGIN

Query Match 5.6%; Score 541; DB 9; Length 554;
Best Local Similarity 100.0%; Pred. No. 4.7e-93;
Matches 541; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0y 6962 gctgtgtatcagagcaacacacttgcagcagctctctagtgaagctcaaaagaa 7021
|||||
Db 14 GCTGTGTATTCAGAGCAACACACTTGTGCTCAGCAGCTCTCTAGAGCTCAAAAAGAA 73
0y 7022 actgagacacacaaacatcgtcgttaataaaggaatctctgcatcacagctctcc 7081
|||||
Db 74 ACTGAAGCACACAAACATTCATGTTAAATTAAGAAATCTCTGTCATCCACGCTCTCC 133
0y 7082 agatcccttgagcagcttgcaaacctgagcagcttaagctaaatcaatcagacacctc 7141
|||||
Db 134 ACATCCTTTGGCAGCTTGCAAACTAGCATGTTAAGCTAAATACCAATCGACAGACCTT 193
0y 7142 ctgaacaaatttaaggtgtataccgaactgctgcaagtcacaaagaatcatagctgac 7201
|||||
Db 194 CTGAACAAATTTAAGGTGTATACCGAATCTGCTGCAAGAAAGATCATAGCTTGATC 253
0y 7202 aaagattatgaagaagacacttgcctgagcaaaaagagacatgagctgacctccaa 7261
254 AAGATTAAGAAAGAGACCTTCTGCTGAGCAAAAGAGCAAGATGAGCTGGACCTCCAA 313
0y 7262 ctgcagctgttgagagcagcagcagaaatgtcggaattctgcattcgaaggaaactcaag 7321
|||||
Db 314 CTGCAGTGTGTTGGAGCAGCAGCGCAAAATAGTGGAATTCGATCTGAGAACTCAAG 373
0y 7322 ttcttgaaatgaattcttgaaatgagctacttttaaaaaagaacaataaattcagagt 7381
|||||
Db 374 TTCTGTGAATTTGAATTCCTGATGAGTACTTTTAAAAAAGCAAAATTAATTCAGAGT 433
0y 7382 gtccagagatgactttcagaaggtgcagaglatcctaataatgaatgagatcaaacctgcaa 7441
|||||
Db 434 GTCCAGGATGACCTTTTCAGAGCTGCAAGTATCTTAATAAACAATGAGTGAACACTGCAG 493
0y 7442 gaagagcttgagacacaaagaagctttagtcagtggttgaggaagatttgagatctgcac 7501
494 GAAGAGCTTTGACACACAAAGAGCTTTATGAGTGCTGTGGAGGAATTTGGAGATCTGCAC 553
0y 7502 g 7502
Db 554 G 554

RESULT 4
B069784 669 bp mRNA linear EST 11-DEC-2001
LOCUS
DEFINITION B069784 NIBB Moch1 normalized Xenopus tailbud library Xenopus
laevis cDNA clone XL054124 5', mRNA sequence.
ACCESSION B069784
VERSION B069784.1 GI:17498144
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 669)
Kitayama,A., Terasata,C., Mochii,M., Ueno,N., Shin-I,T. and Kohara
,Y.

TITLE
JOURNAL
COMMENT
Expressed genes in X. laevis embryo
Unpublished (2001)
Contact: Tadasu Shin-I
Center for Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
source
1. 669
Location/Qualifiers
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="X1054124"
/clone_lib="NIB Mochii normalized Xenopus tailbud
library"
/tissue_type="whole embryo"
/dev_stage="stage 25"

BASE COUNT 253 a 122 c 160 g 134 t
ORIGIN

Query Match 5.5%; Score 531.6; DB 10; Length 669;
Best Local Similarity 88.4%; Pred. No. 3e-91; Matches 589; Conservative 0; Mismatches 74; Indels 3; Gaps 1;

Db 7829 gtcagaagtgtatgcacacaaagtgaagaattgcaattgcaagacaactacta 7888
4 GGCAAGATGATGTCAGACGCAAGAAAGAAAGTTGCAATTTGCCAAAAAATGCTG 63

Qy 7889 agcagaacgcgcgaagctgagcctaagcagatgcagtgtaactaactaaagcaagat 7948
64 AGCAAGACCGGTGAGACCGAGCTTAAGGCAATGGAGTGAACCTGACGTGAAGCAAAAC 123

Qy 7949 aatctgcagctgcacataagaataatagaacctacagaagaattggttgcagaagtgca 8008
124 AGCTTGAGCGCTTCATTAATAAGAAACAAACCTACAGAAAGTTGCCAAAGTCCA 183

Qy 8009 gtaccataaagaagaattgcaaccttaaacctgaagtggttaagattgtaattgga 8068
184 GCACCGTATTAAGAAAGAAATGATTAACCTCAAGAGAGTGGTGAAGATGAATGCA 243

Qy 8069 aaataagttactcaaaagcagaacgaagattgctctacttaagttctgttggaa 8128
244 AAAATAAAGTATTCGAAAGCAACAGACCAAGATGCTTCTTAAGCTTTGTGGAA 303

Qy 8129 gataagaagaagcgttcgtaagttaaagaagacttagcgagcagacagcagacac 8188
304 GATAAGGAGAAGAGCTTCCTAGATTAAAGAAAGAACTTAACAGACGACAGACAC 363

Qy 8189 gatacaacagtttctgttccaaaagatatacaagaattcaacttccctgtgactgt 8248
364 GATCAACAGATTGTGTCCAAAAGATATATCCGAGAGCTCTAGTCCCTTTGACTTGC 423

Qy 8249 ggtgtgtgaagcgttatatgtcagaagcagcagatgttctgtgcagtcggaagcc 8308
424 GGTGTGTGAAGCGGCATAGTCCAGACAGCGCAATGCTCGTTCGACGTGAGAAAGCA 483

Qy 8309 ggcctgaaagagcgttcacatataagaagaataatcatattatatacggactatg 8368
484 ACGTTGGAAGGAGCGTTCATTATAGAGAAATATCTATTCACGGTCTATG 543

Qy 8369 tcaagttctgaa--gatcgaaagaagaacaagcacaattctgacttctcat 8425
544 TCCAAATGTGAAGACGATCTTAAGAAAGAAAGCAAAATAGATGGCACTCTTTCAC 603

Qy 8426 actgatcatatcacagagcgtcgctccacaaaactgaactacagacatggccctgt 8485
|||||

Db 604 ACTGATCATCACACAGAGGTTACTTGCAAAAAGAACTTGACAGCAAGCCCTGTT 663
Qy 8486 actcca 8491
|||||
Db 664 TCTCCA 669

RESULT 5
BM262071/c 583 bp mRNA linear EST 18-DEC-2001
LOCUS da940c12.x3 Blackshear/Soares normalized Xenopus egg library
DEFINITION Xenopus laevis cDNA clone IMAGE:4783871 3', mRNA sequence.
ACCESSION BM262071
VERSION BM262071.1 GI:17925111
KEYWORDS EST
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 583)
Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
, Martin,J., Wylie,T., Underwood,K., Theisling,B., Bowers,Y., Person
,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
Waterston,R. and Wilson,R.
Washu Xenopus EST project, 1999
Contact: Sandy Clifton, Ph.D.
Washu Xenopus EST project, 1999
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Library constructed by Bento Soares and M. Fatima Bonaldo
(University of Iowa). DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
Seq primer: -40up from Gibco
High quality sequence stop: 510.
Location/Qualifiers

REFERENCE
AUTHORS

TITLE
JOURNAL
COMMENT
Unpublished (1999)
Contact: Sandy Clifton, Ph.D.
Washu Xenopus EST project, 1999
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Library constructed by Bento Soares and M. Fatima Bonaldo
(University of Iowa). DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
Seq primer: -40up from Gibco
High quality sequence stop: 510.
Location/Qualifiers

FEATURES
source

1. 583
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:4783871"
/clone_lib="Blackshear/Soares normalized Xenopus egg
library"
/sex="female"
/tissue_type="unfertilized egg"
/cell_type="unfertilized egg"
/dev_stage="unfertilized egg"
/lab_host="DH10B"
/note="Vector: pT73-Pac; Site_1: EcoRI; Site_2: NotI;
PolyA-selected mRNA was prepared from unfertilized Xenopus
laevis eggs. The library was constructed in the vector
pT73-Pac as described in Bonaldo, M.F., Lennon, G. and
Soares, M.B. 'Normalization and subtraction: two
approaches to facilitate gene discovery', Genome Research
6:791-806, 1996. The first strand synthesis was used a
NotI-dT18 primer; double stranded cDNAs were ligated to
EcoRI adapters, digested with NotI, and directionally
cloned into the NotI and EcoRI-digested pT73-Pac vector.
The library contained approximately 7.2 X 10⁵
recombinants, with average insert sizes of 1-1.5 kb."

BASE COUNT 201 a 103 c 180 t

Query Match 5.5%; Score 526.4; DB 10; Length 583;
Best Local Similarity 98.9%; Pred. No. 2.9e-90;
Matches 552; Conservative 0; Mismatches 1; Indels 5; Gaps 2;


```

Db      65  AGAAGAGCTTCTTTGAAAAAGATCATCAGCCTCCAGCAAGCAACTACAGTCAAGAA 6
Oy      1841 gaaga 1845
Db      5  AAAAA 1

RESULT  7
Bg815925 609 bp mRNA linear EST 22-MAY-2001
LOCUS     da075909.v1 Wellcome CRC PCS107 tropicalis egg Silurana tropicalis
DEFINITION CDNA clone IMAGE:4465361 5' similar to TR:042263 042263
            KINESIN-RELATED PROTEIN. ; mRNA sequence.
Bg815925
ACCESSION Bg815925.1 GI:14186905
VERSION    EST.
KEYWORDS   western clawed frog.
SOURCE     Silurana tropicalis
ORGANISM   Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
            Xenopodinae; Silurana.
REFERENCE  1 (bases 1 to 609)
            Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
            Martin,D., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person,
            B., Gibbons,M., Harvey,N., Rilter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
            Washu Xenopus EST project, 1999
            Unpublished (1999)
TITLE      JOURNAL
COMMENT    Contact: Sandy Clifton, Ph.D.
            Washu Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by A. Zorn and J. Mason (Wellcome/CRC Institute
            ). DNA sequencing by: Washington University Genome Sequencing
            Center
            Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/MLN at: info@image.llnl.gov
            High quality sequence stop: 467.
            Location/Qualifiers
                1..609
                /organism="Silurana tropicalis"
                /db_xref="taxon:8364"
                /clone_image="4465361"
                /clone_lib="Wellcome CRC PCS107 tropicalis egg"
                /tissue_type="egg"
                /lab_host="DH10B (phage-resistant)"
                /note="Vector: PCS107; Site.1: NotI; Site.2: EcoRI; cDNAs
                were oligo-dT primed and directionally cloned. Average
                insert size 1.5 kb, range 0.5-4 kb. Library constructed by
                A. Zorn and J. Mason (Wellcome/CRC Institute)."
```

BASE COUNT
ORIGIN

Query Match 4.7%; Score 453.4; DB 10; Length 609;
Best Local Similarity 85.4%; Pred. No. 2; Ce-76;
Matches 505; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

```

Oy      5608 agatgatttaagaagtgtaagaagcatgtcagcagcagagaagataaagttcaggaact 5667
Db      6  AGATGATTTAAGTTGGACTCAAGAGAATTCACAGCAGCAAGAAATAAGTTCAAACAGCT 65
Oy      5668 gacctccagattctctcgtcgcaggaagaatctctcttcgtggaataatcagatgcttta 5727
Db      66 GACCTCCAGATTCTCTACTGGAAGAAAGATCTCTCTTCTAGAAAATCATGATACTAGGA 125
Oy      5728 taatgttgcacctgtgaagaacctctaaagcgaagaagatgactgtaacccagttctaagca 5787
Db      126 TAATGTAAACACCATGAAGAAATCTCTCAGAGAGAGATGATCCTGGACCACTTAAGCA 185
```

```

Oy      5788 acacgtgtctctagaagaattgaaactcttagtctgtctctttaaagaagaagattgcatl 5847
Db      186 ACACGTGTCTCTGGAGAGAGGAAACTTAAGCCTGTGCTTTTAAAGAAAAAGACTTTGCAAT 245
Oy      5848 ggaacacagcagaagaagacaaactgatlgtctgccaggaanaacatagatcacagagaa 5907
Db      246 GGAAACAAGCAGAAAAGAAAAGCTGATCAGCCAGCAAGAAATATATGATCTCATGAGAA 305
Oy      5908 gatataaatatagaagaacagtgtaacttaacaagcaccacatttaagaagaaccttta 5967
Db      306 AATAGCAACTATGAAAGAACACAGCTGCTCAAAATGACACCAATTTTAAAGAAACCTTTTA 365
Oy      5968 tgaagaagagagctctatccagtgtaaggagaacatgctgttgacacagaacaccttag 6027
Db      366 TGAANAAGACGCTTATTCATCTCTAAGAGCAACTGCGCTGACACATGACACCTTAG 425
Oy      6028 ggaacacattgaagagaagaacttgatggtaaaatggggcagggaagagatgaagc 6087
Db      426 GCAACATTTGAAGAGCAAGAACCTGGCATGGGTGAATGAGACAGACAGATGAGAC 485
Oy      6088 tggcaataaagttaatgctctcttaacagaagaatgctctctctagaagaacagatcaatga 6147
Db      486 CGTAATTAAGGTTATACCTCTTACAGAAAATATATCTTTCTTGAAGGAACAGATCAGTGA 545
Oy      6148 aaatgtactactctaaagaagtgtagggcgaaagaagaccttacct 6198
Db      546 AATGCTACTACTCTCTAACAAGTAATGATGAATAATGAAGCCTCAGCCT 596

RESULT  8
AL659719 670 bp mRNA linear EST 13-DEC-2001
LOCUS     AL659719 XGC-neurula Silurana tropicalis cDNA clone TNeu047p06 5',
DEFINITION mRNA sequence.
AL659719
ACCESSION AL659719.1 GI:17673545
VERSION    EST.
KEYWORDS   western clawed frog.
SOURCE     Silurana tropicalis
ORGANISM   Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
            Xenopodinae; Silurana.
REFERENCE  1 (bases 1 to 670)
            Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
            Sanger Centre
            Hinxton, Cambridgeshire, CB10 1SA, UK
            Email: tropes@sanger.ac.uk
            Sanger Xenopus tropicalis EST project 2001
            TROPICALIS_SEQUENCE_ID: TNeu047p06.sp6
            Sequencing primer: SP6
            This sequence is from a Xenopus Gene Collection (XGC) library
            constructed by Aaron M. Zorn.
            Location/Qualifiers
                1..670
                /organism="Silurana tropicalis"
                /db_xref="taxon:8364"
                /clone_image="TNeu047p06"
                /clone_lib="XGC-neurula"
                /dev_stage="neurula"
                /lab_host="Escherichia coli DH10B"
                /note="Vector: PCS107; Site.1: EcoRI; Site.2: NotI; cDNA
                was oligo dT primed from 5' end of poly A+ RNA from neurula.
                EcoRI-NotI cut cDNA was then ligated into PCS107 with
                EcoRI at the 5' end and NotI at the 3' end."
```

FEATURES

source

BASE COUNT
ORIGIN

Query Match

4.5%; Score 433.6; DB 9; Length 670;

Best Local Similarity 84.6%; Pred. No. 1,6e-72;
Matches 567; Conservative 0; Mismatches 74; Indels 29; Gaps 6;

QY 8247 gtagtgtaggaagcggatagtagtcagagcacaacatgctgtgtcgtcagtcggaagaaag 8306
|||||
Db 6 GTGGAGAGGAGGATGGTATATGTGCAGAGCAGCTCATTCGTTGTTCGACATCCG-AAAAG 64
QY 8307 cggcttggaagggcgtgtcacattataagaagaataatcatcattatcacggacta 8366
|||||
Db 65 CACCTTGGAAGGAGGATGGGACATTAATAGAGAAATATCATCTTTATCAGCGACTA 124
QY 8367 tgcgaagtcctga---gagtcgaagaagaagaaggcaaatctgattgctcattcttc 8423
|||||
Db 125 TATCCAACTCTGAAGATGCTTAAGAAAACAAAGCAAAATAGATGTCACCTCTCTC 184
QY 8424 atactgcatcatcacacaagagcgtcgtcacaaaactgaactacagacatgagcctg 8483
|||||
Db 185 ATACTGATCATCATCACAGAGGGTTTCCTCCAAAACTGAACTTACAGACATGGCCCTG 244
QY 8484 ttaactcgaagaagtcgtgaagaatgccaagcctacactagatctccaagaagtcggagt 8543
|||||
Db 245 TTTCTCCAGGAAGGCTGTGAATGCGCAAGCCTGCAGCAGATCTCCAAAGACTCAGAGT 304
QY 8544 cgaagcacaagctgtgtgttcacacaagagtcggaatctacagccaattagtaagt 8603
|||||
Db 305 CCACATATTAACGTGTAGTGTCCACAAACAGGCTGTGAATGTACAGCCAGTATGTAATGT 364
QY 8604 ctccagcagaacccggagatgcat---aaacatactcttcaccaagaagtcgtgactgc 8660
|||||
Db 365 CCCACAGCAAGACCGGGATGCAATAAAAAAGCATGTTCTTCCAGACAGCGGTGAGATGC 424
QY 8661 acaaaaagcgtgcccgtgtcccaacagatcggagatgcccacagcatgtatcc 8720
|||||
Db 425 ATAAAAAGCGTACAGTGTCTCCAAACAGATCTGAGATACATTCAGTCACTGTATCCC 484
QY 8721 ctggcagaagccggactgctcaaaaatc-----taactgaagaagcgt 8762
|||||
Db 485 CAGGCAAGGCTGGACGTGATTAATTCACACAGTTCCTCCACTCAAGACGTAAGGCACTC 544
QY 8763 taltcgacaattgtctctccatgcagaacaagaagaatgacagagaataatcctaattcc 8822
|||||
Db 545 TATTCGGCAATTGG---TCTCCATGCAAAAGCGAAATGTACAGGAATAATCTGATTCGC 601
QY 8823 ctaagagcaaatattgtatgtgaagtcaaaatcgatgcttactgttccatccagttt 8882
|||||
Db 602 CTAAAAAC-CATTTCTTGAATGAGTCAAAATCTTGCTTATGTGTCATCTCACTT 660
QY 8883 tctgaattc 8892
|||||
Db 661 TCGATTAATTC 670

RESULT 9
BI349745 426 bp mRNA linear EST 31-JUL-2001
LOCUS
DEFINITION dae3c09.y3 Blackshear/Soares normalized Xenopus egg library
042263 KINESIN-RELATED PROTEIN.; mRNA sequence.
ACCESSION BI349745
VERSION BI349745.1 GI:15044191
KEYWORDS EST
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 426)
Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D.,
Martin, J., Wylie, T., Underwood, R., Theisling, B., Bowers, Y., Pearson,
B., Gibbons, M., Harvey, N., Riltter, E., Jackson, Y., McCann, R.,
Waterson, R. and Wilson, R.
Washu Xenopus EST project, 1999
JOURNAL Unpublished (1999)

COMMENT

Contact: Sandy Clifton, Ph.D.
Washu Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Library constructed by Bento Soares and M. Patricia Bonaldo
(University of Iowa). DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL et. InfoImage.Llnl.gov
Putative full length read
vector to vector length is 496.

FEATURES

source

1. 426
Location/Qualifiers
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:4678672"
/clone_lib="Blackshear/Soares normalized Xenopus egg
library"
/sex="female"
/issue_type="unfertilized egg"
/cell_type="unfertilized egg"
/dev_stage="unfertilized egg"
/lab_host="DH10B"
/note="Vector: pTR73-Pac; Site.1: EcoRI; Site.2: NotI;
PolyA-selected mRNA was prepared from unfertilized Xenopus
laevis eggs. The library was constructed in the vector
pTR73-Pac as described in Bonaldo, M.F., Lennon, G. and
Soares, M.B. 'Normalization and subtraction: two
approaches to facilitate gene discovery', Genome Research
6:791-806, 1996. The first strand synthesis used a
NotI-dT18 primer; double stranded cDNAs were ligated to
EcoRI adapters, digested with NotI, and directionally
cloned into the NotI and EcoRI-digested pTR73-Pac vector.
The library contained approximately 7.2 x 10⁵
recombinants, with average insert sizes of 1-1.5 kb."

BASE COUNT

145 a 73 c 105 g 103 t

Query Match

Best Local Similarity 99.8%; Pred. No. 8.6e-71;

Matches 425; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 7239 ggcattgagcgtgcagctcaactcagctgcttggagcgcagcaggaagaatgctcgg 7298
|||||
Db 1 GGCATGATGAGCTGCCAGCTCCAGCTGCACTGCACTGTTGGAGCAGCACGCGAATAATGTCGG 60
QY 7299 attctgcatctgaggaactcaagttctgtgaatctgaattctgaattgagttactcttca 7358
|||||
Db 61 ATTCTGATCTGAGGAACCTCAAGTCTGTGAATTTGAATCTTGAAATGAGATTAATCTTTT 120
QY 7359 aaaaagcaaatataattcgaagtgatgctcaggaatgactttcagaagggtgaggtacttctaa 7418
|||||
Db 121 AAAAAGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 180
QY 7419 atcaagtagatcaacaactcagaagaagcttgagcacaagaagaagccttatactgagtg 7478
|||||
Db 181 ATCAAGTAGATCAACACTGCAAGAGAGCTTGAACCAAGAAAGCTTTATGACAGTGT 240
QY 7479 tggagaatttggagatctgcacgtcgaatgtaagaactcagtgaaagcagtcgaacagg 7538
|||||
Db 241 TGGAGCAATTTGGAGATCTGCAGCTGATCTTAAGAAACTCACTGAAGAGCATCAACAG 300
QY 7539 aaaaatggcagctgtcttaccatcagctcttaacaaaagcctaaagcagctgtctc 7598
|||||
Db 301 AAAATAGGCGCATTTGCTTACATACAGCTTTAAACAAAAGGCTTAAGCGAGTGTTC 360
QY 7599 agtcgaataatacaagctgagataaccgtatatactgaaacagcttggagcgaatctgcaag 7658
|||||
Db 361 AGTCGAAATACACGTGAGATTAACCGTATATCTGAACCAAGTGTGAAGCAAAATTCGAAC 420

Dp	362	AAAGTCCTTTGGAAGAAGAACTTGACAAATATCTTGATCTACTGCAGAGATTTGCTGGAGAA	421
Oy	4697	aagctgggaactcytctcaacaacttgaatcctctcagaagaagatygagagctcgtctcta	4756
Dp	422	TAGCTGGAAACTTCTCTCTGGAAATGGCTTATCCCTTTTAAAGAAATGGAACGCTACCTGGTG	481
Oy	4757	ttaaaagatctgacctgaacagagc>tagaaagcttgcctgagtgaaacatctctctaaa	4816
Dp	482	TGCTACGATCTCTCCATCATTAACCTTGCAGACGCTAGCTGAGTGGCTATCTGATTTAAT	541
Oy	4817	gagaatattgacacaaacccctaagaatcattcaatactcaactcaagctcaagtcgagaanaa	4876
Dp	542	GAGATTCTAGACGCGAACCCCTTGACCCCTCCCTCAGAGTCTCATATGCTCAATGCTCAGCTCCCA	601
Oy	4877	cagacagagctcaagcttgctcgaagaatcttcaatcagtcgtctctc	4921
Dp	602	CTGCAAGCGCTACAGTTTGGCTCTTACATCTTTGGTATCACATGGCTGCT	646

RESULT	15
AU117417	
LOCUS	798 bp mRNA linear EST 19-OCT-2000
DEFINITION	AU117417 HEMBA1 Homo sapiens CDNA clone HEMBAl001348 5' mRNA sequence.
ACCESSION	AU117417
VERSION	AU117417.1 GI:10932383
KEYWORDS	EST.
SOURCE	human.

REFERENCE
1 (bases 1 to 798)
AUTHORS
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Negai,T., Sugano,S. and
Isogai,T.
TITLE
HRI human cDNA project
JOURNAL
Unpublished (2000)
COMMENT
Contact: Takao Isogai

1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomcse@ri.co.jp
HHI human cDNA project, 5'- & 3'-end one pass sequencing: Helix
Research Institute: cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

FEATURES	SOURCE	LOCATION/Qualifiers
	1. .798	/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/clone="HEMBA1001348"
		/clone_lib="HEMBA1"
		/tissue_type="Whole embryo, mainly head"
		/dev_stage="embryo, 10 weeks"
		/note="vector: pME185FL3"
BASE COUNT	262 a	137 c 186 g 209 t
ORIGIN		4 other

Query Match	3 %	Score 318	DB 9	Length 798
Best Local Similarity	65.0%	Pred. No. 2e-50		
Matches 506	Conservative 0	Mismatches 257	Indels 15	Gaps 2
QY	86	aagtcacagagacaggggacggtgcgttacccgatttcctccatcatcggtctcaaatg	145	
Db	2	AAGTCCGGCGCGGAGGAGTCTCTGSCCAATTTTCTGGGACACAGTTTCAGCTATAGAG	61	
QY	146	tccgagagagatgcagcttaagtgtgtttgaggttcggccgctatacagagagaca	205	
Db	62	GCGGGAAGAGGACCGTGTGCGCTGCGCCAGTGGCGGCCCTGAAACGACAGAGAA	121	

OY	206	-----ggggggttcbaagccaaccctgcgaatgggaagcctggaaaacaaacacacatttcccaagtc	259
Db	122	TCACCTGGAGAAACTGCCCAAGTTTACTGGGAAACATGCACATAATGTTCATTTATCAAGTT	181
OY	260	gatggacaagaagctcttcaatttcgaatcgctgatttaattcgaatccgaatcaagaagtcgaa	319
Db	182	GATGCAAGTAAATCTTCATTTTGGCTGTGCTCTTTCATGTGTAATGAAAGTACCAAAAAT	241
OY	320	atttcacgaagaatagcaagctaccatcaacacgacgtctgcagggaataatgacaca	379
Db	242	GTGTATGGAAGAAATAGCAGCACATATCATGATTCGCAATACAAAGGCTCAATAGTACT	301
OY	380	ataattgcatacgaacagacacatcttcgaagcaagaacgatacaaatgcatggggaacccaat	439
Db	302	ATATTGGCTTATGACAGACACTGCTTCAGGAAAAACATATACATGATGGGTTCAAGAAAT	351
OY	440	tcaatggygcataatacccccaagcacaacaggaagcttttaaaattatlcaggaatacgg	499
Db	362	CATTGGGAGTTATACCCAGGGCAATTCATGACATTTTCCAAAAAATTAAGAAGTTTCT	421
OY	500	aacgagaagcttctcttaagagcttctctatactggaatttcaatlgaaactctgaaagac	559
Db	422	GATAGGGAATTTCTTTCAGTATCTTACATGGAATAATACAAATGAACCATTTACGAT	481
OY	560	ctacgcctctgatacagaagaagaagccctctgaaatcttcgcaggaatcttaataaagac	619
Db	482	TTACCTCTGGCACACCAAAAATGAACCTTTATTTTCGAGAAGATGTTCATATGAAT	541
OY	620	gttatagtctgcagctcgtactgaagaactctgaaatgtgtccctgaaactgtaatacagctg	679
Db	542	GTGTATGTTGGCTGATCTCCACGAAGAAGTTGTATATATCATACAGAAAATGGCTTGAATGA	601
OY	680	atcaaaaaggcttgaaaaaaacagacatatactgagaagcctaaatgaaatgtaataagtagt	739
Db	602	ATTACAAAAGGAGAAAGACGACGCAATTATGGAAGAACAAAATGAATCAACAGACAGT	661
OY	740	cgttcacatacaataattagaatgtaattgttgaaagccggaacaggaatgaatccacaabaat	799
Db	662	CGNCTCATACCACTTTAGATGATTTTGGAAAGCAGAGGAAGCGTGAACCTTTCAAT	721
OY	800	tcagaagaactctgatgagacctgtaatgtaatactcaactgaaattctgtagaattctgctg	857
Db	722	-----TGTGAAGATCTGGTAAAGNATCCCATTTGAAATTTGGTGTGATCTTGGCG	770

Search completed: October 7, 2002, 16:46:41
Job time: 28073 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 7, 2002, 08:58:48 ; Search time 1309.74 Seconds
(without alignments)
12618.549 Million cell updates/sec

Title: US-09-150-867-2

Sequence: 1 gaattcggagtcggatagg.....ttaaaaaaaaaacggaattc 9626

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_032802:*

- 1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	9626	100.0	9626	20	AA26819	Nucleotide sequence
2	620.4	6.4	8083	23	AA570692	DNA encoding novel
3	620.4	6.4	8303	22	AA58253	Human polynucleoti
c	547.6	5.7	8527	22	AA160039	Human polynucleoti
4	206.6	2.1	581	24	AA561800	Lung small cell ca
5	179.4	1.9	10050	23	AB11011	Drosophila melanog
6	179.4	1.9	3440	22	AA52013	Human polynucleoti
7	141.8	1.5	3932	22	AA52997	Human polynucleoti
8	141.8	1.5	501	24	AA561626	Lung small cell ca
9	138.2	1.4				

c	10	131.2	1.4	12971	23	AB11010	Drosophila melanog
	11	126	1.3	4550	23	AA584836	DNA encoding novel
	12	124.4	1.3	2275	22	AA531140	Human diagnostic a
	13	123.2	1.3	1701	21	AA522377	Human secreted pro
	14	117.6	1.2	3366	21	AA542907	Arabidopsis thalia
	15	107.6	1.1	1950	22	AAH15830	Human cDNA sequenc
	16	106	1.1	814	22	AAH05392	Human cDNA clone (
	17	104.8	1.1	2034	23	AB116733	Drosophila melanog
	18	104.8	1.1	4034	23	AB116732	Drosophila melanog
	19	104	1.1	2352	20	AA587656	Thermomyces lanug1
	20	98.8	1.0	2592	21	AA542909	Arabidopsis thalia
	21	96.4	1.0	7134	23	AB112291	Drosophila melanog
	22	95.8	1.0	1542	22	AA585782	CDNA encoding KSP-
	23	95.8	1.0	1728	22	AA585783	CDNA encoding KSP-
	24	95.8	1.0	3741	22	AAH78013	Nucleotide sequenc
	25	95.8	1.0	3741	22	AA585780	CDNA encoding huma
	26	91.8	1.0	7985	23	AB108391	Drosophila melanog
	27	88	0.9	5093	16	AA099876	KIP1 coding sequen
	28	88	0.9	5093	20	AAV71909	S. cerevisiae KIP1
	29	87.6	0.9	4557	21	AA546311	Arabidopsis thalia
	30	87.6	0.9	4989	21	AA542969	Arabidopsis thalia
	31	86.2	0.9	1107	22	AA585784	CDNA encoding KSP-
	32	85.6	0.9	501	24	AA561778	Lung small cell ca
	33	84.8	0.9	3525	23	AB109775	Drosophila melanog
	34	84.4	0.9	2248	22	AA503053	Human diagnostic a
	35	82	0.9	1230	24	AA523972	Human HSK115 prot
	36	82	0.9	1448	23	AB118979	Human HSK115 prot
	37	82	0.9	4757	24	AA523970	Human kinase in supe
	38	82	0.9	4790	22	AA572653	Human cervical can
	39	81.4	0.8	3399	17	AA705868	Chicken leucocytos
	40	79.8	0.8	2389	16	AA562793	Mouse Khcs cDNA
	41	79.2	0.8	1149	22	AA585781	CDNA encoding KSP-
	42	78	0.8	1128	24	AA523971	Human HSK115 prot
	43	77	0.8	3720	23	AB113337	Drosophila melanog
	44	74.4	0.8	463	21	AA54870	Human ORF425
	45	74	0.8	1407	22	AB117294	Human nervous syst

ALIGNMENTS

RESULT 1	AA26819	standard, DNA; 9626 BP.
ID	AA26819	
XX	AA26819;	
AC	AA26819;	
XX	22-JUN-1999	(first entry)
XX		
DE	Nucleotide sequence of centromere-associated protein-E (CENP-E).	
XX		
KW	CENP-E: centromere-associated protein-E; ATPase activity;	
KW	plus end-directed microtubule motor activity; chromosome congression;	
KW	microtubule binding activity; chromosome movement; mitosis;	
KW	cell proliferation; tumor; metastasis; vascular malfunction;	
KW	inflammatory disease; immune disease; anglogenesis; hypertension;	
KW	restenosis; fungal infection; selective herbicide; fungicide;	
KW	insecticide; plant growth regulator; activator; cancer cell marker; ss.	
XX		
OS	Xenopus sp.	
XX		
PN	WO9913061-A1.	
XX		
PD	18-MAR-1999.	
XX		
PF	10-SEP-1998;	98WO-US19231.
XX		
PR	11-SEP-1997;	97US-0058645.
XX		
PA	(REGC) UNIV CALIFORNIA.	
XX		
PI	Cleveland DM,	Goldstein LSB, Sakowicz R, Wood KW;
XX		

DR WPI, 1999-229233/19.
DR P-PSDB; AAY01632.

XX Centromere-associated protein-E and related nucleic acid

PS Claim 13; Page 67-73; 77pp; English.

XX The present sequence encodes CENP-E (centromere-associated protein-E)
CC of Xenopus. The protein has at least one of plus end-directed microtubule
CC motor activity, ATPase (adenosine triphosphatase) activity and
CC microtubule binding activity. CENP-E is the motor that powers chromosome
CC movement toward microtubule plus ends and is essential for congression
CC of chromosomes during mitosis. Modulators of CENP-E activity are lead
CC therapeutic, bioagricultural and diagnostic agents, e.g. for treatment
CC of unwanted cell proliferation (typical of many examples are tumors and
CC metastases; vascular malfunction; inflammatory and immune diseases;
CC angiogenesis; hypertension; restenosis; and fungal infections), also as
CC plant-protection agents (selective herbicides, fungicides and
CC insecticides) and plant growth regulators or activators for improving
CC yields. CENP-E is also a diagnostic marker for dividing cells, including
CC cancer cells.

XX Sequence 9626 BP; 3540 A; 1692 C; 2104 G; 2290 T; 0 other;

SQ

Query Match 100.0%; Score 9626; DB 20; Length 9626;
Best Local Similarity 100.0%; Pred. No. 0;

Matches 9626; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaatccggagtcgtagtgctgagtcgagaggaattcaactggtatagaagaa 60
Db 1 gaatccggagtcgtagtgctgagtcgagaggaattcaactggtatagaagaa 60
QY 61 ctgaaaccgagcaaaaggagactaaagtacagagacagggagcggtgctacggat 120
Db 61 ctgaaaccgagcaaaaggagactaaagtacagagacagggagcggtgctacggat 120
QY 121 ttccccaatcgtgtctcaaaatgtccgagggagatgcatggttaaagtgtgtgaggt 180
Db 121 ttccccaatcgtgtctcaaaatgtccgagggagatgcatggttaaagtgtgtgaggt 180
QY 181 tcggcgctatatacagagagaaacagggagatcaagccaactgcaatlgagagctgaa 240
Db 181 tcggcgctatatacagagagaaacagggagatcaagccaactgcaatlgagagctgaa 240
QY 241 caacacattcccaagtgtgacgacaagcttccaattcgatcggtatattaatc 300
Db 241 caacacattcccaagtgtgacgacaagcttccaattcgatcggtatattaatc 300
QY 301 tcaagaaatcaacaaatgtaattacaaagaatagcagtaacctatatacgttcagctt 360
Db 301 tcaagaaatcaacaaatgtaattacaaagaatagcagtaacctatatacgttcagctt 360
QY 361 gcaaggatataatgacacaatattgcatagcagacacatcttcaggcaagctaac 420
Db 361 gcaaggatataatgacacaatattgcatagcagacacatcttcaggcaagctaac 420
QY 421 aatgatgggaacacccaattcatcttggtgcaataacccaagccatacaggaaagttctaa 480
Db 421 aatgatgggaacacccaattcatcttggtgcaataacccaagccatacaggaaagttctaa 480
QY 481 aattatcagagatccgaacagagagttctctaaagagttcttataagagattta 540
Db 481 aattatcagagatccgaacagagagttctctaaagagttcttataagagattta 540
QY 541 caatgaatctgtaagagactactgtgtgatgaacagaaagcccttggaattcg 600
Db 541 caatgaatctgtaagagactactgtgtgatgaacagaaagcccttggaattcg 600
QY 601 cgaagatttaataagaacgtgtatgtgtcgtgacctgactgaagaactctgtaagttcc 660
Db 601 cgaagatttaataagaacgtgtatgtgtcgtgacctgactgaagaactctgtaagttcc 660

QY 661 tgaacatgtaatacagtcggtacaaagggtgaaaaaacaagacattatgagagactaa 720
Db 661 tgaacatgtaatacagtcggtacaaagggtgaaaaaacaagacattatgagagactaa 720
QY 721 aatgatgataatagtagtcgttcacatacaataatttagaatgttggaaagccgaga 780
Db 721 aatgatgataatagtagtcgttcacatacaataatttagaatgttggaaagccgaga 780
QY 781 caagaatgatacacaatccagaactgtagtgaagctgtcatgtagtcttcaactgaa 840
Db 781 caagaatgatacacaatccagaactgtagtgaagctgtcatgtagtcttcaactgaa 840
QY 841 ttgtgtagatctgtctgtagcagtgaaagcaagccaactgtagctgaaagtgtagact 900
Db 841 ttgtgtagatctgtctgtagcagtgaaagcaagccaactgtagctgaaagtgtagact 900
QY 901 taaggaagctgtgaataataacccagagctgtttatccttgagacaggttataagaagct 960
Db 901 taaggaagctgtgaataataacccagagctgtttatccttgagacaggttataagaagct 960
QY 961 taagcagcgccaggtctgtggaatttaactacagagacaaactcacagaaatctc 1020
Db 961 taagcagcgccaggtctgtggaatttaactacagagacaaactcacagaaatctc 1020
QY 1021 ccaaaattatgtggagagaaatgctaaacggtataattgtgcaaatlacgcaatttc 1080
Db 1021 ccaaaattatgtggagagaaatgctaaacggtataattgtgcaaatlacgcaatttc 1080
QY 1081 ttgtgtagatcttaagtagacactcagtttgccagtagtccaaacgtgagaaatac 1140
Db 1081 ttgtgtagatcttaagtagacactcagtttgccagtagtccaaacgtgagaaatac 1140
QY 1141 tcccatgttaatgtgaggtccctgtagatgaaagcgtgtgtaaaaggtacagaaagaaat 1200
Db 1141 tcccatgttaatgtgaggtccctgtagatgaaagcgtgtgtaaaaggtacagaaagaaat 1200
QY 1201 ctgtgatttaaaagaaacaatttagagaatttagagttcatcgtctaaacaaagctcaagc 1260
Db 1201 ctgtgatttaaaagaaacaatttagagaatttagagttcatcgtctaaacaaagctcaagc 1260
QY 1261 aatggtctaagaagaagacatacacagttgtagctgaaatcaaaacatacaacaaagag 1320
Db 1261 aatggtctaagaagaagacatacacagttgtagctgaaatcaaaacatacaacaaagag 1320
QY 1321 agaagatagaatagtagcactgacaaatatgtgtgtgtctcatccaaagatctcaaca 1380
Db 1321 agaagatagaatagtagcactgacaaatatgtgtgtgtctcatccaaagatctcaaca 1380
QY 1381 ggaacaaagggtcaaacgaaacgaagaagtacgtggtggtggtggtggtggtggtggtg 1440
Db 1381 ggaacaaagggtcaaacgaaacgaagaagtacgtggtggtggtggtggtggtggtggtg 1440
QY 1441 tttaacgtcttctgt 1500
Db 1441 tttaacgtcttctgt 1500
QY 1501 caagaagcaagagttcctctgacatgcttcattccagaatagtagactcgtgtgtgtac 1560
Db 1501 caagaagcaagagttcctctgacatgcttcattccagaatagtagactcgtgtgtgtac 1560
QY 1561 agagtttctgatttgaatgacgacctccatgtagtgaacgaatgtagatgacaga 1620
Db 1561 agagtttctgatttgaatgacgacctccatgtagtgaacgaatgtagatgacaga 1620
QY 1621 atggaatttagcagatgaatgaatcagagagagaagaatacattcaatcaatgtagat 1680
Db 1621 atggaatttagcagatgaatgaatcagagagagaagaatacattcaatcaatgtagat 1680
QY 1681 agacttggacagatttctgacagtggttcaagttcaatgttcttcttaaggaacacagct 1740
Db 1681 agacttggacagatttctgacagtggttcaagttcaatgttcttcttaaggaacacagct 1740

QY	17141	acaatacctcccgaagactctctgtgtatagtctgaatgacgaaaaagcttcttttgaaaa	1800
Db	1741	acaatacctcccgaagactctgtgtatagtctgaatgacgaaaaagcttcttttgaaaa	1800
QY	1801	agagatcaagaagcttccgaagcaacttaacgcaagaggggagaaaaaggaactgtc	1860
Db	1801	agagatcaagaagcttccgaagcaacttaacgcaagaggggagaaaaaggaactgtc	1860
QY	1861	acaagaactctgaagctcgaagatgaacgaagcttggaagagcgagcttaagtccaagcttaaaaa	1920
Db	1861	acaagaactctgaagctcgaagatgaacgaagcttggaagagcgagcttaagtccaagcttaaaaa	1920
QY	1921	tctagagaatggttaacaaactcgaagagagcatcccataaatgctctgaagctcaacagatgt	1980
Db	1921	tctagagaatggttaacaaactcgaagagagcatcccataaatgctctgaagctcaacagatgt	1980
QY	1981	tgaagaagaagttgttggaagaagaatgtccagtcctcttgagagctctgtttacaatgtatc	2040
Db	1981	tgaagaagaagttgttggaagaagaatgtccagtcctcttgagagctctgtttacaatgtatc	2040
QY	2041	aaacagtgacatccacagatagctctctgtatgtaaacgtccaagcgactcccatgtga	2100
Db	2041	aaacagtgacatccacagatagctctctgtatgtaaacgtccaagcgactcccatgtga	2100
QY	2101	gtgtatagaacacagaaaaatgtcttggaaaaaagaatcgtttgaatttagaagaagtttatga	2160
Db	2101	gtgtatagaacacagaaaaatgtcttggaaaaaagaatcgtttgaatttagaagaagtttatga	2160
QY	2161	aaaccttaacagaagaagaatgtgaataacaaaaactctctgaagcaagaatttatgga	2220
Db	2161	aaaccttaacagaagaagaatgtgaataacaaaaactctctgaagcaagaatttatgga	2220
QY	2221	gagatctccagctctgtgaagaactaaatggcagaagaagaagcaaatgcaactcgagaggaactcggc	2280
Db	2221	gagatctccagctctgtgaagaactaaatggcagaagaagaagcaaatgcaactcgagaggaactcggc	2280
QY	2281	acttaatgcgagataaattctgacaataatatactctagaagaatggaacctctcaaaaagggaaat	2340
Db	2281	acttaatgcgagataaattctgacaataatatactctagaagaatggaacctctcaaaaagggaaat	2340
QY	2341	tgcagatcttggaacgttccactcaaggaagaatcaagaagaacaaatgagtttgaaattctgga	2400
Db	2341	tgcagatcttggaacgttccactcaaggaagaatcaagaagaacaaatgagtttgaaattctgga	2400
QY	2401	gaaggaagaacccaagaaagaagaacgaagccaacaaacttaacgaatgttgcaattgaaga	2460
Db	2401	gaaggaagaacccaagaaagaagaacgaagccaacaaacttaacgaatgttgcaattgaaga	2460
QY	2461	atgtgctggaataatgacgagagaatgtaacaaatcaaaaacttgagagagaatcagaagaactaaac	2520
Db	2461	atgtgctggaataatgacgagagaatgtaacaaatcaaaaacttgagagagaatcagaagaactaaac	2520
QY	2521	aaaaactctgaaagaagcagaatcaacttgcagaatgaagaaacgcgcagataaact	2580
Db	2521	aaaaactctgaaagaagcagaatcaacttgcagaatgaagaaacgcgcagataaact	2580
QY	2581	gcgaaaaaaagtgcgaataatttgatctccggtttccaaatgggtgaatggtggaagaactgtg	2640
Db	2581	gcgaaaaaaagtgcgaataatttgatctccggtttccaaatgggtgaatggtggaagaactgtg	2640
QY	2641	tgaagaagaatcttcaactggaacgaactctctctgagatcgctgaagctggaactcgcgtgc	2700
Db	2641	tgaagaagaatcttcaactggaacgaactctctctgagatcgctgaagctggaactcgcgtgc	2700
QY	2701	tcgaagaagaatgttcttctccctcgaagaatgaaaaatcagaagcttgaaagaagaatctgga	2760
Db	2701	tcgaagaagaatgttcttctccctcgaagaatgaaaaatcagaagcttgaaagaagaatctgga	2760
QY	2761	cacatcaaacctgttaacatacaaaaaaagaagcgtcgctctgttctgttggaagcagcgtgga	2820
Db	2761	cacatcaaacctgttaacatacaaaaaaagaagcgtcgctctgttctgttggaagcagcgtgga	2820
QY	2821	aactggaaaaaatcaaacatacagaagaagaatggaagctgtaattgcaagaagaagaatgttcgaagatgc	2880

Dh	2821	aactgaaataccaacactacaaagaataatggaagctgtgattcgcagaaagagctctgcaaaagtcgc	2880
Qy	2881	ttttaaataagaattaaactaaatctgaatctgacccctccgcagagaaagctccccaagatcttcgc	2940
Dh	2881	tttttaataagaattaaactaaatctgaatctgacccctccgcagagaaagctccccaagatcttcgc	2940
Qy	2941	ttctcgtctgtgaatttagaagaaaaagcttctctgaagctctctcaaaagcagctctgaagaaagcat	3000
Dh	2941	ttctcgtctgtgaatttagaagaaaaagcttctctgaagctctctcaaaagcagctctgaagaaagcat	3000
Qy	3001	ggaaagaaaaaaaatgcctctgtgagaaatgaagtgcactctgcatacagaatacaaaattcttcgc	3060
Dh	3001	ggaaagaaaaaaaatgcctctgtgagaaatgaagtgcactctgcatacagaatacaaaattcttcgc	3060
Qy	3061	aaatgaagttgaatgctctgaaaaaatcagatccagcaagagctctccgaagagatgaatttatt	3120
Dh	3061	aaatgaagttgaatgctctgaaaaaatcagatccagcaagagctctccgaagagatgaatttatt	3120
Qy	3121	aaagcaagaagaagagaaacattctgcatactatacagaacaaacaagaagatattatctgcagga	3180
Dh	3121	aaagcaagaagaagagaaacattctgcatactatacagaacaaacaagaagatattatctgcagga	3180
Qy	3181	gcaatctgaagcagattttaaacaactgcagctcgaagtgcacacacacacagtcacaagctgca	3240
Dh	3181	gcaatctgaagcagattttaaacaactgcagctcgaagtgcacacacacacagtcacaagctgca	3240
Qy	3241	gcaagctctgaagaagaatacttggagataggaaagaatgcagcttgatctcttttgaaagatga	3300
Dh	3241	gcaagctctgaagaagaatacttggagataggaaagaatgcagcttgatctcttttgaaagatga	3300
Qy	3301	tatcagaacaacaagaagctgaagctgcagaagcctctttaaagaaatcgaagaaactctaaagagac	3360
Dh	3301	tatcagaacaacaagaagctgaagctgcagaagcctctttaaagaaatcgaagaaactctaaagagac	3360
Qy	3361	tatgtagtctgtggaagtaaagatctgcgcacacaaacaatgaaacttgaaagaaactataag	3420
Dh	3361	tatgtagtctgtggaagtaaagatctgcgcacacaaacaatgaaacttgaaagaaactataag	3420
Qy	3421	ggataaagaagaacactgctctcaatgaaagaaataatcttttccaagcnaatgcagacataatt	3480
Dh	3421	ggataaagaagaacactgctctcaatgaaagaaataatcttttccaagcnaatgcagacataatt	3480
Qy	3481	tcgcgttaacacctcttccagaactgcctccctccccaataatgaattgaaagggaaactctcca	3540
Dh	3481	tcgcgttaacacctcttccagaactgcctccctccccaataatgaattgaaagggaaactctcca	3540
Qy	3541	agaaccccaatagaataatcaatgaactaccacaatttaataagcctctgtcacaagaaaggaaacaa	3600
Dh	3541	agaaccccaatagaataatcaatgaactaccacaatttaataagcctctgtcacaagaaaggaaacaa	3600
Qy	3601	catatagtctgtctctaaagactgaagaagaaacagctcccaagggcagaagcttatctgattgaa	3660
Dh	3601	catatagtctgtctctaaagactgaagaagaaacagctcccaagggcagaagcttatctgattgaa	3660
Qy	3661	caactaaactttaaagctctcaagcacaagaatattgaaagagctctgtctccccaagccaaa	3720
Dh	3661	caactaaactttaaagctctcaagcacaagaatattgaaagagctctgtctccccaagccaaa	3720
Qy	3721	gcaagactctgaaagaaggaaggtttaaactgctcttctggagctggaactcagaaagggca	3780
Dh	3721	gcaagactctgaaagaaggaaggtttaaactgctcttctggagctggaactcagaaagggca	3780
Qy	3781	ccctaaactgacccacaagctgtctatagaagaattgcagcctcggaagaaatcggaaagctaacga	3840
Dh	3781	ccctaaactgacccacaagctgtctatagaagaattgcagcctcggaagaaatcggaaagctaacga	3840
Qy	3841	aaaactccaacacacttcaagaagaagatgtaaaaaatactactacagaagaagatgagctctcca	3900
Dh	3841	aaaactccaacacacttcaagaagaagatgtaaaaaatactactacagaagaagatgagctctcca	3900
Qy	3901	gaccacactctgaagaccttgaagcagaagcctgaatagccttaaaacaagaactgaattgaaaa	3960
Dh	3901	gaccacactctgaagaccttgaagcagaagcctgaatagccttaaaacaagaactgaattgaaaa	3960

Db 3901 gacccaacttcgaagacctgaagcagagcatgtagctctaagaacacagccttatgtgaaaa 3960
OY 3961 catctagcagctcaatttgaaacacagatgataatgaagctctccaggaagsgctaaagaa 4020
Db 3961 ccttgaagcagctcaatttgaaacacagatgataatgaagctctccaggaagsgctaaagaa 4020
OY 4021 acagaagaacactgtgtctgaagcttctagaacaacagcctttatagatgtctctgaagaatttc 4080
Db 4021 acagaagaacactgtgtctgaagcttctagaacaacagcctttatagatgtctctgaagaatttc 4080
OY 4081 atccaccaaacatgtatgtatgtcttaaacacagaaagatgtcatttgggtgaagttaattc 4140
Db 4081 atccaccaaacatgtatgtatgtcttaaacacagaaagatgtcatttgggtgaagttaattc 4140
OY 4141 gttacaagaatgaaatgtcgtctgtgtgaagaagatgagcttcaaacatctgttgaagcatt 4200
Db 4141 gttacaagaatgaaatgtcgtctgtgtgaagaagatgagcttcaaacatctgttgaagcatt 4200
OY 4201 agtttcagaaactgtgagctactctcgtctcaatgttaaaaactgttggaggagaacacctga 4260
Db 4201 agtttcagaaactgtgagctactctcgtctcaatgttaaaaactgttggaggagaacacctga 4260
OY 4261 aatcacaaaaaaactcaatgtgccttgaaagagatataattggccaatctgaagaagcga 4320
Db 4261 aatcacaaaaaaactcaatgtgccttgaaagagatataattggccaatctgaagaagcga 4320
OY 4321 agtcttgaagtcacatgtcttgagaatctctaaaggagacaacaataagctcaagaagcga 4380
Db 4321 agtcttgaagtcacatgtcttgagaatctctaaaggagacaacaataagctcaagaagcga 4380
OY 4381 agagaatattctctatgaataaataatcaattcagtttagaagaagtggttcaagtgttcaca 4440
Db 4381 agagaatattctctatgaataaataatcaattcagtttagaagaagtggttcaagtgttcaca 4440
OY 4441 gaagctgtgtagatgagaatagaggtccctgaaagacagcagctaaagcgcagagaagaaggtc 4500
Db 4441 gaagctgtgtagatgagaatagaggtccctgaaagacagcagctaaagcgcagagaagaaggtc 4500
OY 4501 ggaataaagaagatagagatattttgaacttctgacaactgtcacaacccaattagtga 4560
Db 4501 ggaataaagaagatagagatattttgaacttctgacaactgtcacaacccaattagtga 4560
OY 4561 gggaaaaattggaactcctatgcaactgtgcaactgtgagagacagcagattgacgcgttc 4620
Db 4561 gggaaaaattggaactcctatgcaactgtgcaactgtgagagacagcagattgacgcgttc 4620
OY 4621 tgaagaataatgagataaaggtccctggagaaaaaacttgaacgaaatcagttattactgga 4680
Db 4621 tgaagaataatgagataaaggtccctggagaaaaaacttgaacgaaatcagttattactgga 4680
OY 4681 aagatttgcagaagaagaagcttggaactgtctgaacaacttgaatcccttcagaaagagat 4740
Db 4681 aagatttgcagaagaagaagcttggaactgtctgaacaacttgaatcccttcagaaagagat 4740
OY 4741 ggaagacgtccgttctattaaagaatgactgtcaacagagaagcttagaagcttgcgcgagga 4800
Db 4741 ggaagacgtccgttctattaaagaatgactgtcaacagagaagcttagaagcttgcgcgagga 4800
OY 4801 aaacatttcttaaaaggaatatttgacaacaacctaaagcatcattcagaatactcaagc 4860
Db 4801 aaacatttcttaaaaggaatatttgacaacaacctaaagcatcattcagaatactcaagc 4860
OY 4861 tcaagctgcagaanaaacacgcgaagcctcagcttgaagaacttccaatcgtctctc 4920
Db 4861 tcaagctgcagaanaaacacgcgaagcctcagcttgaagaacttccaatcgtctctc 4920
OY 4921 tgaacaattgtccaataactcaagaagaagaaacctctgcagatgtgtgcatccttcggga 4980
Db 4921 tgaacaattgtccaataactcaagaagaagaaacctctgcagatgtgtgcatccttcggga 4980
OY 4981 agaaaaagatatgttataacttgaagaattgcatcaaaaaactaagaacaggaanaatt 5040
Db 4981 agaaaaagatatgttataacttgaagaattgcatcaaaaaactaagaacaggaanaatt 5040

OY 5041 actacatgaaagaatgagtaacttgaaacaagctcaaggttgagctaaagtgcgagtggaaca 5100
Db 5041 actacatgaaagaatgagtaacttgaaacaagctcaaggttgagctaaagtgcgagtggaaca 5100
OY 5101 tctgtatgaagagatgataatcgaatcgaagctcctcaacttgaattgtcttcgcagcaagaanaa 5160
Db 5101 tctgtatgaagagagatgataatcgaatcgaagctcctcaacttgaattgtcttcgcagcaagaanaa 5160
OY 5161 tgaatttgacaacaacactctgtcctttaaagaacagatgacgaagtactactcaagaanaa 5220
Db 5161 tgaatttgacaacaacactctgtcctttaaagaacagatgacgaagtactactcaagaanaa 5220
OY 5221 gaaagagctgcacaacaacccaatgaaacacttaacagctgagtggaaccttcaagaagaa 5280
Db 5221 gaaagagctgcacaacaacccaatgaaacacttaacagctgagtggaaccttcaagaagaa 5280
OY 5281 tatgtaatgggttgtaattttaaaaaatgaaagcgcagcgaacaaagaccactaaagacatg 5340
Db 5281 tatgtaatgggttgtaattttaaaaaatgaaagcgcagcgaacaaagaccactaaagacatg 5340
OY 5341 tctgttaaatgggaataaagaacttgagcaaaagccagacagacttcaatgttgagataga 5400
Db 5341 tctgttaaatgggaataaagaacttgagcaaaagccagacagacttcaatgttgagataga 5400
OY 5401 agagcttctgaaagacttaagaagataagagtcagcgcgtggaacactttaaagaatctga 5460
Db 5401 agagcttctgaaagacttaagaagataagagtcagcgcgtggaacactttaaagaatctga 5460
OY 5461 acaaaaagtaatttaataatcgtgaacccaagaatgaaatgttatgttgtaaaatgagagat 5520
Db 5461 acaaaaagtaatttaataatcgtgaacccaagaatgaaatgttatgttgtaaaatgagagat 5520
OY 5521 gaaaaatagccagagagcagctgaattgtctgagaaggagaccagctgcaaacagacctaaggga 5580
Db 5521 gaaaaatagccagagagcagctgaattgtctgagaaggagaccagctgcaaacagacctaaggga 5580
OY 5581 gagtgttgaatgtccatttgaactcaagaatctgaatgtctgaagaagcagctgcaagaagcctaaggga 5640
Db 5581 gagtgttgaatgtccatttgaactcaagaatctgaatgtctgaagaagcagctgcaagaagcctaaggga 5640
OY 5641 gcagcagaagaataaagtctcaaggaactgacctccagatcttcgtctcgtcagagaagaagat 5700
Db 5641 gcagcagaagaataaagtctcaaggaactgacctccagatcttcgtctcgtcagagaagaagat 5700
OY 5701 ctctctctctggaataatagatgtcttataatgtctgcaactgtgaaagaaactctaagcga 5760
Db 5701 ctctctctctggaataatagatgtcttataatgtctgcaactgtgaaagaaactctaagcga 5760
OY 5761 aagaatgacctgaaccagcttaagcaaacacacctgtctcagaagaattgaaactctaagct 5820
Db 5761 aagaatgacctgaaccagcttaagcaaacacacctgtctcagaagaattgaaactctaagct 5820
OY 5821 gtctttaaagaagaagaatttgcatttgaacaacagcagaagaagcaagaactgtcgtgc 5880
Db 5821 gtctttaaagaagaagaatttgcatttgaacaacagcagaagaagcaagaactgtcgtgc 5880
OY 5881 caggaaaacaatagatatacagagaagaatatacaaatatagaagaacagcttaactcaaca 5940
Db 5881 caggaaaacaatagatatacagagaagaatatacaaatatagaagaacagcttaactcaaca 5940
OY 5941 agccacaacatttaaagaagaacacttataagaagaagagcttaccaggttlaagaagaca 6000
Db 5941 agccacaacatttaaagaagaacacttataagaagaagagcttaccaggttlaagaagaca 6000
OY 6001 actggtcttgaacacagaacaccttaaggaaacatttgaagacaaagccttgcgattggg 6060
Db 6001 actggtcttgaacacagaacaccttaaggaaacatttgaagacaaagccttgcgattggg 6060
OY 6061 taaaatggagcagagagagatgaaagctgcgaataaagtataagcttcaagaagaagat 6120
Db 6061 taaaatggagcagagagagatgaaagctgcgaataaagtataagcttcaagaagaagat 6120

[illegible]


```

Db 8281 aatgcttgctgacgctcgaaagacgccttggaagggagcgtgcataattataga 8340
OY 8341 gaataatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 8400
Db 8341 gaataatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 8400
OY 8401 aaaaatcgtatcgtatcgtatcgtatcgtatcgtatcgtatcgtatcgtatcgt 8460
Db 8401 aaaaatcgtatcgtatcgtatcgtatcgtatcgtatcgtatcgtatcgtatcgt 8460
OY 8461 tgaacttaacagacagacagacagacagacagacagacagacagacagacagac 8520
Db 8461 tgaacttaacagacagacagacagacagacagacagacagacagacagacagac 8520
OY 8521 aggaatcccaagaagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 8580
Db 8521 aggaatcccaagaagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 8580
OY 8581 aatcacaagcgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 8640
Db 8581 aatcacaagcgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 8640
OY 8641 tccaagcaaggttgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 8700
Db 8641 tccaagcaaggttgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 8700
OY 8701 caccacagcgtgacatccctcgtgagcgtgagcgtgagcgtgagcgtgagcgtgag 8760
Db 8701 caccacagcgtgacatccctcgtgagcgtgagcgtgagcgtgagcgtgagcgtgag 8760
OY 8761 gtatcgaacatctgctcctcgtgagcgtgagcgtgagcgtgagcgtgagcgtgag 8820
Db 8761 gtatcgaacatctgctcctcgtgagcgtgagcgtgagcgtgagcgtgagcgtgag 8820
OY 8821 ccctaagagcaaatattatgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgag 8880
Db 8821 ccctaagagcaaatattatgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgag 8880
OY 8881 ttctgtaaatcctcaagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgag 8940
Db 8881 ttctgtaaatcctcaagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgag 8940
OY 8941 aagtcagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 9000
Db 8941 aagtcagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 9000
OY 9001 atcctagatccctgtaacatcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 9060
Db 9001 atcctagatccctgtaacatcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 9060
OY 9061 ctgtgtagaagaacaacatcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgag 9120
Db 9061 ctgtgtagaagaacaacatcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgag 9120
OY 9121 aagtgctctttagtagtaacatcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 9180
Db 9121 aagtgctctttagtagtaacatcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 9180
OY 9181 caatacgtcgaacgtcctttagtagtaacatcgtgagcgtgagcgtgagcgtgagcgtgagcgtgag 9240
Db 9181 caatacgtcgaacgtcctttagtagtaacatcgtgagcgtgagcgtgagcgtgagcgtgagcgtgag 9240
OY 9241 ctgtgactcataattggtcgtcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 9300
Db 9241 ctgtgactcataattggtcgtcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 9300
OY 9301 aaaaatgacatttgacgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 9360
Db 9301 aaaaatgacatttgacgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 9360
OY 9361 agcacatagtggtgacgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 9420
Db 9361 agcacatagtggtgacgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 9420

```

```

OY 9421 gaaatactacaaagatagggcaaaagtgtttctcactagtgtaaacagatagtggtt 9480
Db 9421 gaaatactacaaagatagggcaaaagtgtttctcactagtgtaaacagatagtggtt 9480
OY 9481 aacttaattatctccttaataaattctatataatgagcgtcattagtgactaat 9540
Db 9481 aacttaattatctccttaataaattctatataatgagcgtcattagtgactaat 9540
OY 9541 ggtattcgcgcctcttgaacaccttttaatttactcgttaataagcgcctgat 9600
Db 9541 ggtattcgcgcctcttgaacaccttttaatttactcgttaataagcgcctgat 9600
OY 9601 tttaattaaaaaaacggaattc 9626
Db 9601 tttaattaaaaaaacggaattc 9626

RESULT 2
ID AAS70692 standard; cDNA; 8083 BP.
XX
AC AAS70692;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #6496.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN W0200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001MO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX
P-PSDB; ABG06505.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 1, SEQ ID No 6496; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human

```

CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 8083 BP; 3384 A; 1234 C; 1642 G; 1823 T; 0 other;

Query Match 6.4%; Score 620.4; DB 23; Length 8083;
Best Local Similarity 65.5%; Pred. No. 4.9e-131;
Matches 1001; Conservative 0; Mismatches 491; Indels 36; Gaps 5;

QY 69 gccgcgaagaaggactaaagtacagagagagcggtgttcgtaaccgatttccac 128
DB 17 gcgccctgtgagccctggaagctgcgcgcgagggctccgcaattctgttgagccag 76
QY 129 taatcgtctcaaaatgtcccgagggagatgcagttaaagtgtgtgtgaggttcgcgcgc 188
DB 77 ttcaagctgatatgattgtgagagagagagcggtgcgtctgcgtgcgaggtgcgcgcgc 136
QY 189 ttatacagaagaaaca-----ggggatcaagccaacctgtgaaatggaagcttgaaaca 242
DB 137 tgaacagcagagaaatcaactctgtgagaaactgcccaagtttacttgaaacttgacaata 196
QY 243 aacacattcccaagttgtgtgagacaagttcttaattctgatacgtgtatttaattcc 302
DB 197 atgccttatacagttcgtatgagaaagtaaatccctcaattctgacgtgtcttcaatgta 256
QY 303 acgaatcaacaagctcaaatcttccaagaataagcaatlacctatacagatcagctttgc 362
DB 257 atgaaactacaaatgtgtatgaaagaataagcagcacacatcatcgattctgcacatc 316
QY 363 agggatataatgtgcacaataattgtcatcacgacagacatcttcaggaagctgtacaaa 422
DB 317 aaggctacataatgtctataattgtcctaagacagactcttcaaggaataaacatatacca 376
QY 423 tgatggaaacacccaattctatgtggcataataaccccaacccaagaggaagttttaaaa 482
DB 377 tgatgggttcagaaagatcatcttcggagcttatcccgaggaatctacatgtcattccaaa 436
QY 483 ttatcagaagatlacgaaacagagagttcttcttaagagttcttctaataagagattaaa 542
DB 437 aaattcaagaagttctcctgtataggaattctctcttaagtgatcttaccatggaataata 496
QY 543 atgaaactgtgaaagactactgtgtgtatgacagaaagaaagcccttggaaattcgcg 602
DB 497 atgaaacccaattacagattactctgtgcacccaataaaatgaaaccttaattatctcag 556
QY 603 aggattttaagaaacgtgtatgtgtgcacctgacgtgaagaacttgaattgttctcg 662
DB 557 aagatgtcacaatggaatgtgtatgtgtgcacccaagaaagaaagctgtatataatcag 616
QY 663 aacatgtcaatcagttgatacaaaagggtgaaanaaaacagacattctggaagactaaa 722
DB 617 aaatgtccttgaaatgtgattacaagggagaaagagcgacttatgtggaacaaa 676
QY 723 tgaatgatcatcgtcgtctacacatacatattgaatgattgttgaagcggagaca 782
DB 677 tgaataaagaagacgctgtcttcacatacactcttagagatttctggaagaagagaga 736
QY 783 gaaatgatcccaaatctagaagaactgtgagatgtcatagttatgattccttgaatt 842
DB 737 aggtgtgaaccttcaat-----tgtgaaggaactgttcaaggtatcccatgtgaatt 787
QY 843 tggtagatcttctgtagcagtgaaagagcaagccaactgtgaaagggtgtgagacta 902
DB 788 tggtagtcttgcaggcagtgaaagagcgtctcaaaacagcgctgtagaggtgtgcgctca 847
QY 903 aggaaggtctgacaatcacaacgcagctgttataccttvgacaaggttataaagaacta 962
DB 848 aggaaggtctgataataatctgaagcttattatttggacaagtgatcaagaacta 907
QY 963 ggcagcgccaggtctgtgtgattataataactacagaaacgcaaacaccagatcttcc 1022

DB 908 gtgatgacaaagtctgtgttccataataatcogagatagcaagttacaacgaaatcttc 967
QY 1023 aaattcatctggagaggaataatgtcaaaacggttataatttgcacaatlaacgcgaattct 1082
DB 968 agaatctcttggaggaataatccaagaacgattatcttcgacaattatctccagattctt 1027
QY 1083 ctgattagactcaagtacaacttcagtttgcagatcttcgccaacatctgagaatactc 1142
DB 1028 ttgatgaactcttaactgtctcctcagtttgcagatcttgcagattctgataatataagaaatc 1087
QY 1143 cccattgttaatgagtcctgagatgaaagcgtgtgttaaaaggtcagaagaagaatctc 1202
DB 1088 cttaagttaatgtgagatcaacacgtgaaagctccctcgaaaggtatgaaagaataaa 1147
QY 1203 tggatttaagaagaacaattagagaatttgaagatcagtcgtctgaaacaaagctcagaagca 1262
DB 1148 tggattcttaaaaaacaattagag-----gaggttctctttagagacggtgcagcgca 1201
QY 1263 tggctaaagaaggtcattacagattctgactgaaatcaaaacacacacaaagagaag 1322
DB 1202 tggaaanaagaccaaatttgcgccaaacttctgaaagaaagaatttcttcgaaggtacaga 1261
QY 1323 aagatagaatgtgcactgtgacaatatgtgtgtg---cttcatcccaagaaatcacaac 1379
DB 1262 atgagaagaattgaaacttaacacagagatcgtgtgacccctcttccccaagttgcac 1321
QY 1380 aggaacaaagggtcacaacgaagaagttacgtgtggcgccaggaanaaatcaaaata 1439
DB 1322 aggaataaaggtcaaaacgaagaagttactcgtgtcccttgcgaataatacaaaa 1381
QY 1440 gttatactcttctgtgttcttcgactttgatatgtctatccagattactctggaatttta 1499
DB 1382 tgaagaa-----ctcaactatgcagataatltaaatatacaaaacaaatataa 1429
QY 1500 gcaagaaggcaaaagtctcgtgacatgccttcacattccagaatattgatacgtctgtga 1559
DB 1430 caacaaaacacataagcttctcatabaatttatataagagaagaattgtatgaatcgtctgt 1489
QY 1560 cagaatttcttgatttgatgacgcct 1587
DB 1490 cagagctgtatgttccagtaacacct 1517

RESULT 3
AA158253
ID AA158253 standard; cDNA; 8503 BP.
XX
AC AA158253;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 456.
XX
DE Human; neurotic; immunosuppressant; cytostatic; gene therapy; cancer;
XX
KW Human; neurotic; immunosuppressant; cytostatic; gene therapy; cancer;
KW Alzheimer's; Parkinson's disease; Huntington's disease; hemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotoxic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW Leukemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000MO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.

RESULT 4
AA160039/c
ID AA160039 standard; cDNA; 8527 BP.
XX
AC AA160039;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 4028.
XX
KW Human; nocitropic; immunosuppressant; cytostatic; gene therapy; cancer;
peripheral nervous system; neuropathy; central nervous system; CNS;
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PE 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QN, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR P-PSDB: AAM40883.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 4028; 10078bp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AA442213) with nocitropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 8527 BP; 1982 A; 1715 C; 1302 G; 3525 T; 3 other:

Query Match 5.7%; Score 547.6; DB 22; Length 8527;
Best Local Similarity 64.2%; Pred. No. 2e-114;
Matches 988; Conservative 0; Mismatches 504; Indels 48; Gaps 9;

QY 69 gccgcacaaaggaagtaagtgacagagagagcggtgtcgtatccgattccccc 128
DB 8511 GCGGCTGTGAGCCCTGAATGCTCCGCCGCGAGGGTCTGGCATTTTGTGGGAC 8452
QY 129 taatcgctcacaatgtccgaagagatgcagtttaagtgtgtgtgaggttcggcgcg 188
DB 8451 TTCAGCCTGATAGGATGGCGGAGGAGACCCTGGCCGTCTCGTCGAGTGGCCGC 8392
QY 189 ttatcagaagagacaa-----gggatcgaagcacaacctgcagtgaagctggaaca 242
DB 8391 TGAACAGACAGAGAAGATCACTTGAGAAACCTGCCCAAGTTTACTGGAAAACCAATTA 8332
QY 243 acacattcccaagtgtatgagacaagtcttcaattcgtgtgtgttcaattcc 302
DB 8331 ATGTCATTTATCCAGTGAAGTAATCACTTCAATTTTGTGCTGTCTTCAAGTA 8272
QY 303 acgaatacaagaagtaaatctaacaaag-aatagacagtaactcaatgaatcagcttg 361
DB 8271 ATGAACCTCCCAAAATGTGTATGAAGCCAAATGACAGACCAATCATGATTCGCCNTA 8212
QY 362 caggagataatgycacaatattg-catacggacagac-alcitcaggaagacgtaca 419
DB 8211 CAAGCTACAAATGATATATTTGCCCTATGACAGACTGCTTCAGGAAAACATATA 8152
QY 420 caatgttggaacacacaatctctggcctaatacccaagcacaacaggaagtttc- 478
DB 8151 CCATGATGGCTTCAGAAATCATTTTGGAGTATTAACCCAGGCGCAATTTGATGACATTA 8092
QY 479 -----aaattatcagagataccgaagagagttctctcaaggtctctata 530
DB 8091 TTTCACAAAATTTTAAGAAAGTTTCTCTGATAGGGAATTTCTTACGATCTTACA 8032
QY 531 tggagattcaatgaactgtgaagaccctactctgtgtgtagacagaagaaagacct 550
DB 8031 TCGAAATATATCAATGAACCATTAACAGATTACTGTGGACATCAAAAATGAAACCTT 7972
QY 591 tggaaattcgagagatttaataagaacgtgtatgtctgcaccgcgaagaaacttg 650
DB 7971 TAATTTATCGAGAGATGTCAAATAGCAATGTATTTGCTGTATTCACAGAAAGTTTG 7912
QY 651 taatgtctcctgacatgttaatacagtgatcaaaaggtgaaagaaagacatcatg 710
DB 7911 TATATCATATCAGAAATGCGCTTTGAATGATTCAAAGGAGAAAGACGACCATTTAG 7852
QY 711 gagagactaaatgaatgacatagtagtgcgtcacaataatttgaatgtgtgtg 770
DB 7851 GAGAAACAAATGATCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7792
QY 771 aaagccgagacagaatgatccacaataatcagaagaactgtgtagagctgtcatg 830
DB 7791 AAAGCAG 7741
QY 831 ctactctgaattgtgtagatctgtcggcagtgaaagagcaagcacaactgagagcgaag 890
DB 7740 CCATTTGAATTTGTTGATCTTGTGACAGCGAGTAAAGAGCTGCTCAAAAGGGGCTGAC 7681
QY 891 gttgagactaaaggaagctgtcaacatacaacgcagctgtgttcaatcctctgagcagtt 950
DB 7680 GTGTGGGGCTCAAG 7621
QY 951 ttaagaagcttagcagcggcagcgtgtgtgatttataactacagagagacaactca 1010
DB 7620 TCAGAAACCTTACTGATGAGCAAGTTGTGCTTCAATTAATTTGAGAGAGAGAGAGAG 7561
QY 1011 ccagaattctcaaaatctaatctgaggaagaaatgtcaaaacgtttatcaattgcaactta 1070
DB 7560 CACGAATTTCTTCAGAAATCTTGGGAGAGAAATCCAAAGACAGCTATTTATCTGACAAATTA 7501
QY 1071 cgcgaattccttcttgaagagctcctaaagtaactcagttctgcaagtaactgcaaacag 1130
DB 7500 CTCACATATCTTTTGTATGAAGCTTACTGCTCTCCAGTTTGGCAATGACGTAATTA 7441

QY 1131 tgagaatactcccatgttaatgagtcctcgatgaagcgtgtgctaaaggtaca 1190
DB 7440 TGAGAAATACCTTATAGTTATATAGATACCACTGATGAGAGCTCTCTGAAAAGGTATA 7381
QY 1191 gaaaggaatcttgatattaaagaacaattagagaattagagtcacgtctgaacaa 1250
DB 7380 GAAAAGAAATATATGATCTTAAAAAACAATTAGAG-----GAGGTTCTTTAGAGACGC 7327
QY 1251 aagctcaagcaatggtcaagaagaagacatacagctgtgctagctgaatcaacaactac 1310
DB 7326 GGGCTCAGGCAATGAAAAGCAATGGCCCACTTTTGAGAAAAGAAATTTGCTTC 7267
QY 1311 acaaaagagagaagatagatatgacctgtgaacaatatgtgtgtgt---ctcatccc 1367
DB 7266 AGAAAGTACAGAAATGAGAAAATTGAAAACGAGATGCGGAGACCTCTTCTTCC 7207
QY 1368 aagaatcacagagaccagaaggtcaaaagaaaagaagatgagtgagcgccagaa 1427
DB 7206 TCACGTTGCAACAGAAATTAAGGCTTAAAGAAAACGAAGAGTTACTGTGCTTGCGCA 7147
QY 1428 aaatccaaatagttacatgctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1487
DB 7146 AAATTAACAAATGAGAA-----CTCAAACTATGCAATCAATTAATATAC 7099
QY 1488 ctggaattttagcaagaagagcaagattctctgaacatgcttcaattccagaatgatg 1547
DB 7098 CAACAATATTAACAAACAAACACATTAAGCTTCTTAAATTTATTAACAGAAATTAAG 7039
QY 1548 actctgtgttacaagagttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1587
DB 7038 AATCTGTCTGTCAAGCTGTGATGTTTTCAGTAACACTCT 6999

RESULT 5

AA61800
ID AA61800 standard; cDNA; 581 BP.
XX
AC AA61800;
XX
DT 29-JAN-2002 (first entry)
XX
DE Lung small cell carcinoma antigen, cDNA #341.
XX
KW Human; cytosolic; antitumor; lung small cell cancer antigen;
XX
OS tumor; lung cancer; ss.
XX
FN Homo sapiens.
XX
PN WO200177168-A2.
XX
PD 18-OCT-2001.
XX
PF 11-APR-2001; 2001WO-US11859.
XX
PR 11-APR-2000; 2000US-196780P.
PR 21-JUN-2000; 2000US-213361P.
PR 01-SEP-2000; 2000US-229763P.
PR 05-SEP-2000; 2000US-230629P.
PR 14-SEP-2000; 2000US-232565P.
PR 19-DEC-2000; 2000US-257037P.
PR 08-JAN-2001; 2001US-260796P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Lodes MJ, Wang T, Mohamath R, Indirias CY;
XX
DR WPI; 2002-010896/01.
XX
XX
PT Lung tumour polynucleotide and polypeptides useful in therapy and
PT diagnosis of cancer especially lung cancer
XX
PS Claim 1; Page 247; 295pp; English.
XX

CC The invention relates to novel isolated lung small cell cancer antigen
CC polynucleotides (I) and polypeptides (II) used in a method of detecting
CC cancer in a patient. The method is optionally performed by
CC utilising oligonucleotides (III), where the biological sample
CC from the patient is contacted with (III), detecting the amount of
CC polynucleotide hybridised to (III) in the sample and comparing the
CC amount of polynucleotide to a predetermined cut-off value and thereby
CC determining cancer in a patient. (I), (II) or antigen-presenting cells
CC expressing (II) is useful for stimulating and/or expanding T cells
CC specific for a tumour protein. The method comprises contacting T cells
CC with one of the components under conditions to permit the stimulation
CC and/or expansion of the cells. A composition comprising (I) is useful for
CC stimulating an immune response in a patient and for inhibiting the
CC development of a cancer especially lung cancer and for inhibiting the
CC isolated T cell population is useful for removing tumour cells from the
CC biological sample and for inhibiting the development of cancer in a
CC patient. AA61460-AA61874 represent novel human lung small cell
CC cancer antigen coding sequences of the invention.
XX

Sequence 581 BP; 205 A; 101 C; 130 G; 145 T; 0 other;

Query Match 2.1%; Score 206.6; DB 24; Length 581;
Best Local Similarity 68.4%; Pred. No. 3.9e-37;

Matches 333; Conservative 0; Mismatches 144; Indels 10; Gaps 3;

QY 955 gaagcttagcgagcgagcgtgtgtattatataaactacagacacacacacag 1014
DB 72 gaaacttagtgatgaaagtggtgttcaataaattatcgcagatgagcaagtaacag 131
QY 1015 aattcccaaatcattgtgagaggaatgctaaacggtatattgtgcaattacgcc 1074
DB 132 aattcccaaatcattgtgagaggaatgctaaacggtatattgtgcaattacgcc 191
QY 1075 agttctttagatgagacacttaagtaacttcagttgtccagatgacgcacaacatgtgag 1134
DB 192 agtactctttagatgagacacttcagttgtccagatgacgcacaacatgtgag 251
QY 1135 aaatctcccatgttaataggtccctgagatgagagcgtgtgtcaaaaaggtacagaa 1194
DB 252 gaataccctcttagatgagatgagatgagatgagatgagatgagatgagatgagatgag 311
QY 1195 ggaatctttagatgagatgagatgagatgagatgagatgagatgagatgagatgag 1254
DB 312 agaaatagatgagatgagatgagatgagatgagatgagatgagatgagatgagatgag 365
QY 1255 tcaagcaatggtcaagaagaagacacacagttgctagctgaatcaacaactacacaa 1314
DB 366 tcagcaatggtcaagaagaagacacacagttgctagctgaatcaacaactacacaa 424
QY 1315 agagagagaagatagatagatgagatgagatgagatgagatgagatgagatgagatgag 1371
DB 425 agtacaagatgagatgagatgagatgagatgagatgagatgagatgagatgagatgag 484
QY 1372 atctcaagagagacaaaggtcaaacgaaacgaaagatgagatgagatgagatgagatgag 1431
DB 485 gtgcaccccgagatgagatgagatgagatgagatgagatgagatgagatgagatgag 544
QY 1432 ccaaat 1438
DB 545 accaaat 551

RESULT 6

AB11011
ID AB11011 standard; cDNA; 10050 BP.
XX
AC AB11011;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 27515.
XX

CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SRO ID NO 2110 (AKR52581), 2111 (AKR52582) and 3666
CC (AKR80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX Sequence 3932 BP; 1312 A; 771 C; 913 G; 936 T; 0 other;

Query Match	1.5%	Score 141.8;	DB 22;	Length 3932;
Best Local Similarity	52.8%;	Pred. No. 5.5e-22;		
Matches 504; Conservative	0;	Mismatches 397;	Indels 54;	Gaps 7.

OY	263	gggcaagctcttcaatttcgatctgtatatttaattctcaagatcaacagttcaatt	322
Db	558	ggctccagctcttatgcattttgtatctgggttccagttcaagacactccaagaagatgt	617
OY	323	taccagaagaatacgaattaccatcatcatgaatcagcttttcagggatataatggccacata	382
Db	618	tataatgacctgtccaagaagatgtgttaagaatgtacttggaaaggatataatggaaacata	677
OY	383	ttgcatacggagaagacatcttcagggcaagagctaacacaatgtatggaacac-----	434
Db	678	ttgcatacgtacaaacatcctctcgtgggaagacacacaaatggagggttaacttcgat	737
OY	435	-caaatcatctgggcataataaccaccaagccatacaggaagttttaaattatcagaag	493
Db	738	ccgaagaaggcatgggaattatctccaagaataatgtccaagatatttttaattatcttacc	797
OY	494	at---accgacaagagatcttcttccaagagttctctatcatatgagatcttacaatgaact	550
Db	798	atgagtgaataatttgagattcatatgaatgttcaatcttgaataatatttgataag	857
OY	551	gtgaagaacctactgtgtgtatgacagaagaagaagcccttggaaatccgcgagatttc	610
Db	858	ataagggacctgttgaatgtt-----tcaaaagaccaactcttcagttcatatgaagacaa	911
OY	611	aatgaagaactgtatgtttgtctgaacccgtactggaagaactgtatgttcttcgaatga	670
Db	912	aaccggaattccctatgttaaaagggtgtcacagaagcgtttgtatgtcagtcocgaatgaatc	971
OY	671	atacagttgacatcaaaaagggtgaaaaaaaacagacattatggagagactaaatgaatgat	730
Db	972	atgatacccatagatgaagaagaataatccacaagacatgtagcagttatacaattgatga	1031
OY	731	catatgtagtgcgtccactcaatacaatttagaattgattggttgaagcccggaacgaatgat	790
Db	1032	catatgctctctagsgctccagttatattcttattaaagt-----caacaagaagaa	1081
OY	791	cccaacaattcagaagaactgtgacatggagctgtcatatgtaatcctcacttgaattgtgaat	850
Db	1082	cacacaagaagcaacaagaagtgtg-----gaaacttatctgtgtat	1127
OY	851	ctctgtcgcagtgaagaagacgaaccgaactgtgaagctggaaggtgtgaagactaaagaagc	910
Db	1128	ttagctgtgtatgtgaaaagggttagtaaaactctggagcgtgaagggtgtgtctgtatgaaact	1187
OY	911	tgcacatcaaccgcagctgtgttatcccttggacaaggttattaagaagcttggacaagc	970
Db	1188	aaaaaacctcaacaagctcaacttcttgctcttggaaatgtaattcttcgctttgcttgaagg-	1246
OY	971	caggctgtgtatattaaactcatcagagaagcagaacatccacagaattctccaanaatc	1030
Db	1247	-----tagtcatatgttccatcatctcgagatgttaaatgtacaagaatccctccaagaatca	1301
OY	1031	ttgggaagaagaatgtctaaagaagttataatttgcacaattagccaggttcttgttatgt	1090
Db	1302	ttaggvtgcacactgtlaaacactacttgaatttgaatttgcgtctccatccatcaatacgaatg	1361
OY	1091	acttaa-----gtaacttcagtttgcgcagttactgcacaacaatgvtgaagaattccc	1144

Db 1362 tctgaacaaatctacactcttattgycgaagggccaaacaaatgaagacacgct 1421

Dy 1145 catgttaatggcgccctgagtgaaagagctgtctaaanaaggtacacgaagggaa 1199

Dd 1422 tctgtcaatgttgagcttaactgcagacaagctggaaaagaagatgtaaaaaagaa 1476

RESULT	9
AAS61626	
ID	AAS61626 standard; cDNA; 501 BP

AA	AAS61626;
AC	
XX	
DT	29-JAN-2002 (first entry)

DE Lung small cell carcinoma antigen, cDNA #167.

Human; cytostatic; antitumour; lung small cell cancer antigen;
tumour; lung cancer; ss.

OS	Homo sapiens.
XX	
PN	WO200177168-A2.

PD	18-OCT-2001.
XX	
PF	11-APR-2001: 2001WO-US11859.

XX 11-APR-2000; 2000US-196780P
XX 21-JUN-2000; 2000US-213361P
PR 01-SEP-2000; 2000US-228763P
PR 05-SEP-2000; 2000US-230629P
PR 14-SEP-2000; 2000US-233656P
PR 19-DEC-2000; 2000US-257037P
PR 08-JAN-2001; 2001US-260796P
PA (CORI.) CORIXA CORP.

xx Lodes MJ, Wang T, Mohamath R, Indirias CY;
 xx
 xx WPI, 2002-010896/01.
 xx
 xx Lung tumour polynucleotide and polypeptides useful in therapy and
 xx PT diagnosis of cancer especially lung cancer -
 xx PT
 xx Claim 1; Page 180; 295pp: English.
 xx PS

The invention relates to novel isolated lung small cell cancer antigen polynucleotides (I) and polypeptides (II) used in a method of detecting cancer in a patient. The method is optionally performed by utilizing oligonucleotides (III), where the biological sample from the patient is contacted with (III), detecting the amount of polynucleotide hybridised to (III) in the sample and comparing the amount of polynucleotide to a predetermined cut-off value and thereby determining cancer in a patient. (I), (II) or antigen-presenting cells expressing (II) is useful for stimulating and/or expanding T cells specific for a tumour protein. The method comprises contacting T cells with one of the components under conditions to permit the stimulation and/or expansion of the cells. A composition comprising (I) is useful for stimulating an immune response in a patient and for inhibiting the development of a cancer especially lung cancer in a patient. An isolated T cell population is useful for removing tumour cells from the biological sample and for inhibiting the development of cancer in a patient. AAS61460-AAS61874 represent novel human lung small cell cancer antigen coding sequences of the invention.

SQ Sequence 501 BP; 205 A; 67 C; 101 G; 128 T; 0 other;

Query Match	1.48;	Score 138.2;	DB 24;	Length 501;
Best Local Similarity	57.78;	Pred. NO. 1.5e-21;		
Matches 266; Conservative	0;	Mismatches 193;	Indels 2;	Gaps 1;


```
OY 2844 aaatggaagctgatttgcagaagaagttgcagaagtgctttaaagatnaactactta 2903
Db 1 aaatgcaagttgattcgtgaagaagaatatactatgctttaaagataaacaactca 60
OY 2904 atgacctcttgcaggaagaagttccccaagattgcttctcgtgttgaaatagaanaa 2963
Db 61 cctccctataatgagcaaaagttccaaaagattgctctcgtgaatttgaaataggaa 120
OY 2964 aggtctcgaattctcaaacagctgtgaaaga--ttggaagaaaaaatgcttga 3021
Db 121 agattcgtatcttcagaagaactaataaagaagttgaagaagaatgaagcttcg 180
OY 3022 gaatgaagtgactgctgctatcaagaataaattttgccaatgaatgaatgcttga 3081
Db 181 ggaagaagtcatttgcgttcgaagattgaattcttactccttcgaagtaagaagctcga 240
OY 3082 aaatcagatcagcaagcttctgaaagataatgttataagcaagaaggaacaattc 3141
Db 241 gaaagaagatacaagaacaattctgaagctccatataacatcagaaaaagataaatt 300
OY 3142 tgcatacataaagcaacaagaagattatcatgagcaggaatctgagagaatttaca 3201
Db 301 gttctcgaagtagttatcataagaagtagaagttcaagcttacttgaaagaattggaa 360
OY 3202 actgactgacgaagtgacacacacaaagtcacgaagtcagacgactgaaagaactactt 3261
Db 361 aacaaagaagctacactacacagctcgaattataaagaagcagatcaagaattcca 420
OY 3262 ggaagatgaagaataatgatgatcttcttttgaagtata 3302
Db 421 aaattcaaaacccctcatatgagcttgcagcaaaagtata 461

RESULT 10
ABLI1010/c
ID ABLI1010 standard: cDNA: 12971 BP.
XX
AC ABLI1010;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 27512.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI: 2001-656860/75.
XX
DR P-PSDB: ABB66907.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1: SEQ ID NO 27512; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
```

CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI1617-ABLI30511), expressed DNA
CC sequences (ABLI01840-ABLI16175) and the encoded proteins
CC (ABBI737-ABBI7072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX

Sequence 12971 BP: 3384 A: 3000 C: 2690 G: 3897 T: 0 other:

Query Match 1.4%; Score 131.2; DB 23; Length 12971;
Best Local Similarity 51.1%; Pred. NO. 2.4e-19;
Matches 445; Conservative 0; Mismatches 378; Indels 47; Gaps 4;

```
OY 403 ttcaagcaagctacacatgatgatggaacccaattcatggtcacaatacccgaagc 462
Db 11249 TTCAGGCAAGACGTACACCATATATGCGCAGACAGAAATCCGGCTGATGCTTAGC 11190
OY 463 catacagaagttttaaattatcaggagagataccgaacagagagttcttcaagat 522
Db 11189 CCGCAAGGAGATCTTCAACAGATCTCCAGTGAACGAGGGGACTTTCTGCTGGCT 11130
OY 523 tcttatatgagatttacaatgaactctgtgaagaactactgtgtgatacagaagaa 582
Db 11129 GGGCTACATCGACGTCTCAACGAGAAAGATCTACATCTGCTG-----AACACGAGAA 11076
OY 583 gaagcccttgaatctcgagagatttataagaagctgtatgttgcgtccgactga 642
Db 11075 TCAGGACCTTCAAGATCCATGAGTCCGGAAATGGGATTTGTATGT-----GAACCTCGAGGA 11019
OY 643 agaacttgtaatggttctcgtgaacatgtaatacaglygatacaaaaaggttgaanaaacag 702
Db 11018 GTGCATCATTAACACGACGAGTTGACCTCCGCTGTATGCTTGGGCAATATAGAGCG 10959
OY 703 aattatgagagactaaatgaatgatcatagtagtcgttcaatcataatttagaat 762
Db 10958 CACCGTGGCGAGAGAACATGAAAGAGCGGTCCAGCGCTTGGCATCATATTAAAGAT 10899
OY 763 gatgtltgaa-----agcagagacaagaat 787
Db 10898 AGTAAGTCCAGCTGCTATTTCTTCCGATTAATCTCAACATGTAATCCACAGATCATC 10839
OY 788 gatccacaattcaagaactgtgatg---agctgcatggtatctcaactgaattg 844
Db 10838 GAGTCCCGAAGTGGACACACAGCATGACGACGACGATTCGATTCGAAAGCGTTGAACCTG 10779
OY 845 gtagacttctgacgtgaaagagcaagcaactggaactggaagtgatggaacttaag 904
Db 10778 GTGATCTGGCTGGGTGAGAGCGGCGGACCAACGGGCTCTGAGAGAGCGCCCTTCAA 10719
OY 905 gaaggtgcaacaatcaacgcagctgttatactcttgaagaagttataagaagcttagc 964
Db 10718 GAGGCGCGTCAATATAACAGAGCTTACTTTCTCGATGACTTTCAAAAGATCTGTG 10659
OY 965 gaagcgcagctggtgattataaactacagagacagaactccagaaattctcaa 1024
Db 10658 GAATAACCGGACAAATGATTTACCACTACCGGATTCAGAGCTGACGGCATTTGCGAG 10599
OY 1025 aattcattggagaaatgtaanaacggtataatttgacaattacagcttcttctt 1084
Db 10598 GCTTCTTTGGGGGCAATGATTCATCATCATCTATTTCGACATTAACCAATCATGATCATG 10539
OY 1085 gatgaacttcaatgaactcaagtttgcagctactgccaacaatgtgaaatactccc 1144
Db 10538 GAAGAGTCCAGTCTACCTCTAGTTGTCGACGCGCAAGAAATTTGCAATTAAGCCC 10479
OY 1145 catgtaatgaggtcctgtagatgaagcgttcaaaaagttacagaagaagaattctg 1204
Db 10478 CAAGTCAACGAGATGTGATCTGAGCCACAAATGATCAAGCGCTGAGGAGAAATTAAG 10419
```

[illegible]


```
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0158294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 1.2% Score 117.6; DB 21; Length 3366;
Best Local Similarity 54.2%; Pred. No. 1.7e-16;
Matches 412; Conservative 0; Mismatches 309; Indels 39; Gaps 7;

```
OY 471 aagttttaaattatcttcgagagacccgaacagagagttcttcttaagattcttata 530
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 830 aggaagtgaagtgtcttaacgggactccaggaaggaggttctgtctcggtttcatatc 889
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 531 tggagattacaatgaaactgtgaaagacctatgtgtatgacagaaagaagccct 590
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 830 ttgaatatatacatgagtgtgataatgatttactg-----gatccaacagccaaact 943
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 591 tggaaatccgagagatttataagaacggtatgtgtcgtacacctgacttgagaacttg 650
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 944 taagtgttagagaggtatcccgagg--caactatgttgaaggtatcaaggagaagtgtg 1000
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 651 taatgttccctaaacatgaatacagtgatcaaaagaagtgaaaaaaacagacattatg 710
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1001 ttgtgtccctcgtgcacatgcattcatctgtcagcgttgagggaagaacatcgtcatgtg 1060
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
OY 711 gagagactaaatgaatgatcatgagtgtgtctacatatcaaatattagaatgtgttg 770
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1061 gtccaataattcatcttctgtgtgacgcaggaagccaacatatcttaacatgacgtgtg 1120
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 771 aaagccgagacagaatgatccccaatattcaagaactgtgacggagctgtcatgtat 830
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1121 aaagtatgtg-----ccacgtgagatgatatgtgtgag---gtatctcttt 1162
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 831 ctcaacttgaatttgttagatctgtctgtgcagtgaaagcaagccaactgtgagctgaag 890
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1163 ctcaactcaatttgaattgacttgcgtgatctgtgaggttcga--aaactgaanaacgtg 1219
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 891 gtgtgagacttaaggaagcgtgcaacatcaaccgcgcgtgtgttatcccttggacgtta 950
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1220 gatttgaggaggaagggtttcatatcatacaagaaggtcttcaactcttggaaactgtga 1279
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 951 ttaagaagcttaagcagccgcagcgtgtgtgatttataactacagaagaacgaactca 1010
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1280 ttggaanaacttaagtgaagggttaaggcaactcacattcca---tatcgtgaacttaagtca 1336
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1011 ccagaattctccaaaattcaatttggagaagatgtctaaacggtataaatttgcacattta 1070
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1337 ctctgtcttcgcacatcttcaatttaagtgtgtcatatgtgtcgtcatatgtacaattta 1396
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1071 cgtcagtttc-----tttgtatgagacttcaagtcaacttgcactgttgcagatctgcca 1124
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1397 ctctctgtccgcagcagtaagtgaagaactcatatacaacattgaagtgttgcagagggcaga 1456
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1125 aaacatgtgaaatactcccatcgttaatgaggtccctgtgatgtgaaggtgtgtctaaata 1184
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1457 agagtataagaataatataatgcttccacgcacatcaagattataagatggaagatcaatcaaga 1516
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1185 gttacagaagaagaaatcttggattttaaagaaacaattaga 1224
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1517 aatatcaaaagagaatcttcaaccctcaaaactgaactgtga 1556
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

RESULT 15
AAH15830
ID AAH15830 standard; cDNA; 1950 BP.
XX
AC AAH15830;
XX
DT 26-JUN-2001 (first entry)
XX
XX
DE Human cDNA sequence SEQ ID NO:14319.
XX
KW Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT primer sets for synthesizing polynucleotides, particularly the 5602
full-length cDNAs defined in the specification, and for the detection

THIS PAGE BLANK (USPTO)